Maximum *a posteriori* estimators in ℓ^p are well-defined for diagonal Gaussian priors

Ilja Klebanov[1,*] and Philipp Wacker[2]

[1] Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany
[2] School of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand

E-mail: klebanov@zedat.fu-berlin.de

Abstract

We prove that maximum *a posteriori* estimators are well-defined for diagonal Gaussian priors μ on ℓ^p under common assumptions on the potential Φ. Further, we show connections to the Onsager–Machlup functional and provide a corrected and strongly simplified proof in the Hilbert space case $p = 2$, previously established by Dashti *et al* (2013 *Inverse Problems* **29** 095017); Kretschmann (2019 *PhD Thesis*). These corrections do not generalize to the setting $1 \leqslant p < \infty$, which requires a novel convexification result for the difference between the Cameron–Martin norm and the *p*-norm.

Keywords: inverse problems, maximum *a posteriori* estimator, Onsager–Machlup functional, small ball probabilities, sequence spaces, Gaussian measures

(Some figures may appear in colour only in the online journal)

1. Introduction

Let $(X, \|\cdot\|_X)$, be a separable Banach space and μ a centered and non-degenerate Gaussian (prior) probability measure on X. We are motivated by the inverse problem of inferring the unknown parameter $u \in X$ via noisy measurements

$$y = G(u) + \varepsilon, \tag{1.1}$$

---

* Author to whom any correspondence should be addressed.

where $G : X \to \mathbb{R}^d$ is a (possibly nonlinear) measurement operator and ε is measurement noise, typically assumed to be independent of u. The Bayesian approach to solving such inverse problems (Stuart 2010) is to combine prior knowledge given by μ with the data-dependent likelihood into the posterior distribution μ^y given by

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu}(u) = Z^{-1} \cdot \exp(-\Phi(u)). \tag{1.2}$$

Here, the so-called potential $\Phi : X \to \mathbb{R}$ depends on G and the statistical structure of the measurement noise ε, while $Z := \int_X \exp(-\Phi(u))\mu(\mathrm{d}u)$ is simply the normalization constant, which is well defined under suitable conditions on Φ (see assumption 2.1 later on). If, for example, the measurement noise is distributed according to a centered Gaussian measure on $\mathbb{R}^d$, $\varepsilon \sim N(0,\Gamma)$ with symmetric and positive definite covariance matrix $\Gamma \in \mathbb{R}^{d \times d}$, then $\Phi(u) = \frac{1}{2}\|\Gamma^{-1/2}(y - G(u))\|^2$, but we will use general formulation (equation (1.2)) as the starting point for our considerations. For an overview of the Bayesian approach to inverse problems and a discussion of its well-posedness we refer to (Stuart 2010) and the references therein.

Our focus lies on the analysis of the so-called 'maximum *a posteriori* (MAP) estimator' or 'mode', i.e. the summary of the posterior μ^y in the form of a single point $u_{\mathrm{MAP}} \in X$. In the finite-dimensional setting $X = \mathbb{R}^k$, if μ^y has a continuous Lebesgue density ρ^y, MAP estimators are simply defined as the parameter of highest posterior density, $u_{\mathrm{MAP}} = \mathrm{argmax}_{u \in \mathbb{R}^k} \rho^y(u)$ (note that such maximizers may not be unique or fail to exist).

Unfortunately, this definition does not generalize to measures without a continuous Lebesgue density, in particular it cannot cover infinite-dimensional settings, where there is no equivalent of the Lebesgue measure.

For this reason Dashti *et al* (2013, definition 3.1) suggested to define MAP estimators as 'maximizers of infinitesimally small ball (posterior) mass', see definition 1.3 below. To simplify notation, we first introduce the following shorthand for the ratios of ball masses:

Notation 1.1. For a separable metric space X and a probability measure ν on X, we denote the open ball of radius $\delta > 0$ centered at $x \in X$ by $B_\delta(x)$. Further, for $w, z \in X$ with $\nu(B_\delta(z)) > 0$, we set

$$\mathfrak{R}_\nu^\delta(w,z) := \frac{\nu(B_\delta(w))}{\nu(B_\delta(z))}, \qquad \mathfrak{R}_\nu^\delta(w, \sup) := \frac{\nu(B_\delta(w))}{\sup_{z \in X} \nu(B_\delta(z))}.$$

Similarly, we set $\mathfrak{R}_\nu^\delta(\sup, w) := \mathfrak{R}_\nu^\delta(w, \sup)^{-1}$ whenever $\nu(B_\delta(w)) \neq 0$.

Remark 1.2. Note that $\sup_{z \in X} \nu(B_\delta(z)) > 0$ follows from the separability of X: assume that $(z_n)_{n \in \mathbb{N}}$ is dense in X, $\delta > 0$ and $\nu(B_\delta(z_n)) = 0$ for each $n \in \mathbb{N}$. Then $\nu(X) \leqslant \sum_{n \in \mathbb{N}} \nu(B_\delta(z_n)) = 0$ (since $X \subseteq \bigcup_{n \in \mathbb{N}} B_\delta(z_n)$) and ν could not be a probability measure.

We work with the following rather general definition of MAP estimators:

Definition 1.3 (Ayanbayev *et al* 2021a, definition 3.6). Let X be a separable metric space and ν be a probability measure on X. A *strong mode* for ν is any $z \in X$ satisfying

$$\lim_{\delta \searrow 0} \mathfrak{R}_\nu^\delta(z, \sup) = 1. \tag{1.3}$$

If $\nu = \mu^y$ is a Bayesian posterior measure given by (1.2), then we call any strong mode a *MAP estimator*.

Other sources, especially from the physics community, see e.g. (Dürr and Bach 1978), (informally) define the MAP estimator as the minimizer of the so-called Onsager–Machlup (OM) functional, which can be thought of as a generalization of the negative posterior log-density (Dashti *et al* 2013):

Definition 1.4. Let μ be a Gaussian (prior) measure on a separable Banach space X with Cameron–Martin space $(E, |\cdot|_E)$ and $\Phi \colon X \to \mathbb{R}$ be such that $\exp(-\Phi)$ is μ-integrable. We define the *Onsager-Machlup (OM) functional* $I \colon E \to \mathbb{R}$ corresponding to μ^y given by (1.2) by

$$I(u) := \Phi(u) + \frac{1}{2}|u|_E^2. \tag{1.4}$$

The connection between OM minimizers and MAP estimators is non-trivial in general separable Banach spaces[3]. Natural questions arising in this context are

- whether (or under which conditions) MAP estimators exist and
- whether MAP estimators can equivalently be characterized as minimizers of the OM functional.

One fundamental ingredient, and the most direct reason why small-ball probabilities are related to the functional I, is the following theorem about the OM functional:

Theorem 1.5 (Dashti *et al* 2013, theorem 3.2). *Let assumption 2.1 hold. Then for $z_1, z_2 \in E$,*

$$\lim_{\delta \searrow 0} \mathfrak{R}_{\mu^y}^{\delta}(z_1, z_2) = \exp(I(z_2) - I(z_1)).$$

However, theorem 1.5 does not yield the full answer regarding the connection of MAP estimators and OM minimizers—not only is it restricted to elements of the Cameron–Martin space E, also it only provides *pairwise* comparisons of two points $z_1, z_2 \in E$, while MAP estimators require consideration of the ratio $\mathfrak{R}_{\mu^y}^{\delta}(z_1, \mathrm{sup})$ and its limit as $\delta \searrow 0$.

Remark 1.6. Note that I amounts to a Tikhonov–Phillips regularization of the misfit functional Φ, so the results in this manuscript are also to be understood in the context of regularized optimization.

Dashti *et al* (2013) discussed, for the first time, the existence of MAP estimators as well as their connection to minimizers of the OM functional, in the specific setting of a Bayesian inverse problem of type (1.1). More precisely, they claim to prove the following statements for every separable Banach space X under assumption 2.1 below (Dashti *et al* 2013, theorem 3.5):

(I) Let $z^\delta = \mathrm{argmax}_{z \in X} \mu^y(B_\delta(z))$. There exists a subsequence of $(z^\delta)_{\delta > 0}$ that converges strongly in X to some element $\bar{z} \in E$.

(II) The limit $\bar{z}$ is a MAP estimator of μ^y (this proves existence of such an object) and it is a minimizer of the OM functional.

However, while the ideas of Dashti *et al* (2013) are groundbreaking, their proof of the above statements, as well as the corrections provided by Kretschmann (2019), rely on techniques that hold in separable Hilbert spaces rather than separable Banach spaces, see section 1.1.

Further, neither Dashti *et al* (2013) nor Kretschmann (2019) show the existence of the δ-ball maximizers z^δ above, which are the central objects in their proofs. It turns out that the existence of z^δ is a highly non-trivial issue and has recently been discussed by Lambley and

---

[3] Note that (Dashti *et al* 2013, theorem 3.2), restated as theorem 1.5 below, only gives partial answers, since only pairwise comparisons of points lying in E are made, while (Ayanbayev *et al* 2021a, proposition 4.1) makes the connection between OM minimizers and weak modes (rather than strong modes, which correspond to MAP estimators) under different assumptions.

Sullivan (2022), who proved their existence for certain measures (including posteriors arising from non-degenerate Gaussian priors on ℓ^p) and gave counterexamples for others.

Our approach relies on asymptotic maximizers in the following sense, which are guaranteed to exist by the definition of the supremum (in fact, even for arbitrary families $(\varepsilon^\delta)_{\delta>0}$ in $(0,1)$).

Definition 1.7. Let X be a separable metric space and ν be a probability measure on X. A family $(\zeta^\delta)_{\delta>0} \subset X$ is called an *asymptotic maximizing family (AMF)* for ν, if there exists a family $(\varepsilon^\delta)_{\delta>0}$ in $(0,1)$ such that $\varepsilon^\delta \searrow 0$ as $\delta \searrow 0$ and, for each $\delta > 0$,

$$\mathfrak{R}_\nu^\delta(\zeta^\delta, \sup) > 1 - \varepsilon^\delta. \tag{1.5}$$

Lemma 1.8. *For any separable metric space X and any probability measure ν on X, there exists an AMF for ν. Further, if $\bar{z}$ is a MAP estimator for ν, then the constant family $(\bar{z})_{\delta>0}$ forms an AMF for ν.*

Proof. This follows directly from the definition of the supremum (in fact, for *any* family $(\varepsilon^\delta)_{\delta>0}$ a corresponding asymptotic maximizing family (AMF) can be found) and definitions 1.3 and 1.7. □

The corresponding statements to (I)–(II) are given in conjecture 2.3. Note that we strengthened those statements by stating the equivalence of MAP estimators, minimizers of the OM functional and limit points of AMFs. Especially the latter cannot be expected for the δ-ball maximizers z^δ, even when they exist and are unique, since it is easy to construct MAP estimators that are not limit points of $(z^\delta)_{\delta>0}$ as $\delta \searrow 0$, even for continuous measures on $\mathbb{R}^1$. Apart from their guaranteed existence, this is yet another advantage of working with AMFs $(\zeta^\delta)_{\delta>0}$ rather than with $(z^\delta)_{\delta>0}$.

1.1. Why this paper is necessary

The contribution of this paper is twofold:

1. Remedy the shortcomings of previous work on the existence of MAP estimators mentioned above and listed in detail below, resulting in a corrected and strongly simplified proof of the existence of MAP estimators in the Hilbert space setting (theorem 2.4, proven in section 3);
2. Generalize the corresponding result from Hilbert spaces to sequence spaces $X = \ell^p$, $1 \leqslant p < \infty$, of pth-power summable sequences and diagonal[4] and nondegenerate Gaussian prior measures, proven in section 4). For this purpose, we develop a novel and non-trivial convexification argument for the difference between the Cameron–Martin norm $|\cdot|_E$ and the ambient space norm $\|\cdot\|_X$ in proposition 4.6.

The shortcomings of previous work on the existence of MAP estimators include:

- The crucial object in the proofs of (Dashti *et al* 2013),

$$z^\delta = \operatorname{argmax}_{z \in X} \mu^y(B_\delta(z)),$$

---

[4] By 'diagonal' we mean that $\mu = \otimes_{k \in \mathbb{N}} \mathcal{N}(0, \sigma_k^2)$ has a diagonal covariance structure with respect to the canonical basis, while 'nondegenerate' refers to the fact that the eigenvalues of the covariance operator are strictly positive, $\sigma_k^2 > 0$ for $k \in \mathbb{N}$. Note that Gaussian measures on separable Hilbert spaces can always be diagonalized in this sense by choosing an orthonormal eigenbasis of the covariance operator, see notation 3.1, hence our results constitute a genuine generalization of the Hilbert space case.

is defined without a proof of its existence. This is a highly non-trivial issue which was not fixed by the corrections in Kretschmann (2019). In (Lambley and Sullivan 2022, example 4.8), the authors construct a probability measure on a separable metric space without such δ-ball maximizers z^δ, but prove in (Lambley and Sullivan 2022, corollary 4.10) that such maximizers exist for posteriors arising from non-degenerate Gaussian priors on ℓ^p.

- Specific Hilbert space properties are used in Banach spaces, in particular, the proof of (Dashti *et al* 2013, theorem 3.5) relies heavily on the existence of an orthogonal basis of the Cameron–Martin space which satisfies $\|x\|_X^2 = \sum_{n\in\mathbb{N}} x_n^2$ for $x \in X$, where x_n are the coordinates of x in that basis.

- While the defining property of a MAP estimator $z \in X$ is given by

$$\lim_{\delta \searrow 0} \mathfrak{R}_{\mu^y}^\delta(z, \sup) = 1,$$

the proof of (Dashti *et al* 2013, theorem 3.5) considers this limit only for a specific null sequence $(\delta_m)_{m\in\mathbb{N}}$. This is hidden in their notation, where, for simplicity, they adopt the notation $(z^\delta)_{\delta>0}$ for subsequences—a rather typical abuse of notation which is illegitimate in this specific case, since different null sequences $(\delta_m)_{m\in\mathbb{N}}$ can yield different candidates for MAP estimators.

- While Dashti *et al* (2013, lemma 3.9) is stated for $\bar{z} = 0$, it is later applied to more general $\bar{z} \in X$. In Banach spaces, validity of this substitution is equivalent to tacit assumption of the Radon–Riesz property, which only holds for a strict subset of separable Banach spaces (and excludes the paradigmatic case $X = \ell^1$).

- The proof of (Dashti *et al* 2013, corollary 3.10) relies on MAP estimators being limit points of $(z^\delta)_{\delta>0}$. However, only the reverse implication had been discussed, and, in fact, this implication is incorrect even when z^δ, $\delta > 0$, is guaranteed to exist, as can be easily seen from the following simple example of a bimodal distribution on $\mathbb{R}^1$: let $0 < \sigma < 1$ and μ^y have Lebesgue density $\rho^y(x) \propto \exp(-(x-1)^2/2)\chi_{\mathbb{R}^+} + \exp(-(x+1)^2/(2\cdot\sigma^2))\chi_{\mathbb{R}^-}$. Then $z^\delta = 1$ for all $\delta < \frac{1}{2}$, but both $x = \pm 1$ are true MAP estimators. For this purpose, we work with AMFs introduced in definition 1.7, the limit points of which we show to coincide with MAP estimators.

Conjecture 2.3 in general separable Banach spaces and general Gaussian measures remains unsolved and is an extremely intricate issue. The 'skeleton' of our proofs is provided in theorem 2.8, where the main steps are shown under suitable conditions (while proving those conditions in specific settings typically requires a lot of work). This establishes a framework for proving conjecture 2.3 in other Banach spaces, thereby paving the road for future research on this topic.

1.2. Related work

The definition of strong modes by Dashti *et al* (2013) has sparked a series of papers with variations on this concept, most notably generalized strong modes Clason *et al* (2019), weak modes (Helin and Burger 2015). Agapiou *et al* (2018) studied the MAP estimator for Bayesian inversion with sparsity-promoting Besov priors. The connection between weak and strong modes was further explored in Lie and Sullivan (2018), and Ayanbayev *et al* (2021a, 2021b) discussed stability and convergence of global weak modes using Γ-convergence. Recently, Lambley and Sullivan (2022) presented a perspective on modes via order theory.

1.3. Structure of this manuscript

Section 2 describes the common framework along which the well-definedness of MAP estimators can be proven in all cases considered (Hilbert space and $X = \ell^p$) and, possibly, further separable Banach spaces. Sections 3 and 4 apply this framework in order to prove well-definedness of the MAP estimator in the Hilbert space and ℓ^p case, respectively.

2. Existence of maximum *a posteriori* estimators

This section covers all the main results mentioned in the introduction. Throughout the paper, we will make the following general assumptions:

Assumption 2.1. *Let $(X, \|\cdot\|_X)$ be a separable Banach space, which we call the ambient space, and μ be a non-degenerate centered Gaussian (prior) probability measure on X. Let $(E, |\cdot|_E)$ denote the corresponding Cameron–Martin space and μ^y be the (posterior) probability measure on X given by (1.2), where the potential $\Phi : X \to \mathbb{R}$ satisfies the following conditions:*

(a) Φ is globally bounded from below, i.e. there exists $M \in \mathbb{R}$ such that for all $u \in X$,

$$\Phi(u) \geqslant M.$$

(b) Φ is locally bounded from above, i.e. for every $r > 0$ there exists $K(r) > 0$ such that for all $u \in X$ with $\|u\|_X < r$ we have

$$\Phi(u) \leqslant K(r).$$

(c) Φ is locally Lipschitz continuous, i.e. for every $r > 0$ there exists $L(r) > 0$ such that for all $u_1, u_2 \in X$ with $\|u_1\|_X, \|u_2\|_X \leqslant r$ we have

$$|\Phi(u_1) - \Phi(u_2)| \leqslant L(r) \|u_1 - u_2\|_X.$$

Purely for convenience, we assume that $\Phi(0) = 0$. This can be easily achieved by subtracting $\Phi(0)$ from Φ and incorporating the resulting additional prefactor into the normalization constant Z in (1.2).

Remark 2.2. Conditions (a)–(c) are identical to (Dashti *et al* 2013, assumption 2.1), except that (a) is slightly stronger: (Dashti *et al* 2013) initially assume the weaker inequality $\Phi(u) \geqslant M - \varepsilon \|u\|_X^2$ for every $\varepsilon > 0$, but also make the additional assumption of global boundedness from below (in the sense of (a) in assumption 2.1) in their main theorem 3.5. This assumption is usually not too restrictive as our condition (a) still covers most practical Bayesian inverse problems, since Φ is typically even non-negative (cf introduction). Further, the non-degeneracy of μ together with the above conditions guarantees that the ratios $\mathfrak{R}_\mu^\delta(w, z)$ and $\mathfrak{R}_{\mu^y}^\delta(w, z)$ etc are always well-defined. Given the assumption $\Phi(0) = 0$, condition (b) is an implication of (c), but we keep the conditions separated for didactical reasons and comparability to previous papers.

First, let us restate the result in (Dashti *et al* 2013, theorem 3.5) as a conjecture, since their proof is only (partially, due to unclear existence of δ-ball maximizing centers z^δ) correct in Hilbert spaces and the Banach space version remains an open problem:

Conjecture 2.3. Let assumption 2.1 hold. Then:

(a) The following statements are equivalent:
    (i) $\bar{z}$ is an X-strong limit point as $\delta \to 0$ of some AMF for μ^y.[5]
    (ii) $\bar{z} \in E$ and $\bar{z}$ minimizes the OM functional.
    (iii) $\bar{z}$ is a MAP estimator.
(b) There exists at least one MAP estimator.

The main goal of this paper is to provide proofs of conjecture 2.3 in the special cases where

- X is a separable Hilbert space (theorem 2.4), where we correct and strongly simplify the proofs initially proposed by (Dashti *et al* 2013) and worked out in detail in the PhD Thesis of Kretschmann (2019), or
- $X = \ell^p$ with $p \in [1, \infty)$ and $\mu = \otimes_{k \in \mathbb{N}} \mathcal{N}(0, \sigma_k^2)$ is a diagonal Gaussian measure on X (theorem 2.5), which is an entirely new result.

Theorem 2.4. *Let assumption 2.1 hold. Then conjecture 2.3 holds for any separable Hilbert space $X = \mathscr{H}$.*

Proof. See section 3.   □

Theorem 2.5. *Let assumption 2.1 hold. Then conjecture 2.3 holds for $X = \ell^p$, $p \in [1, \infty)$, and any diagonal Gaussian (prior) measure $\mu = \otimes_k \mathcal{N}(0, \sigma_k^2)$ on X.*

Proof. See section 4.   □

2.1. Proof strategy

In order to prove theorems 2.4 and 2.5, we proceed along the following seven steps, where $(\zeta^\delta)_{\delta>0}$ is an arbitrary AMF for μ^y and $(\delta_m)_{m \in \mathbb{N}}$ denotes an arbitrary null sequence. This is a rather general approach and can be followed to prove conjecture 2.3 for further classes of Banach spaces.

  (i) Show that $(\zeta^{\delta_m})_{m \in \mathbb{N}}$ is bounded.
  (ii) Extract a weakly convergent subsequence of $(\zeta^{\delta_m})_{m \in \mathbb{N}}$, which, for simplicity, we denote by the same symbol, with weak limit $\bar{z} \in X$.
  (iii) Prove that $\bar{z}$ lies in the Cameron–Martin space E.
  (iv) Show that the convergence is, in fact, strong: $\|\zeta^{\delta_m} - \bar{z}\|_X \to 0$ as $m \to \infty$.
  (v) Infer that any limit point $\bar{z}$ of an AMF (not just the one obtained in (ii)–(iv)) is a MAP estimator, proving its existence.
  (vi) Prove that any MAP estimator minimizes the OM functional and is a limit point of some AMF.
  (vii) Show that any OM minimizer is also a MAP estimator.

An illustration how this proof strategy fits within the context of conjecture 2.3 can be found in figure 1.

The proof of (i), (iii) and (iv) is highly non-trivial and relies on the following idea: First, we prove that, under assumption 2.1, the fraction $\mathfrak{R}_\mu^{\delta_m}(\zeta^{\delta_m}, 0)$ is bounded away from 0, meaning

---

[5] I.e., there exists a sequence $(\delta_n)_{n \in \mathbb{N}}$ with $\delta_n \searrow 0$ such that $\|\zeta^{\delta_n} - \bar{z}\|_X \to 0$ as $n \to \infty$.
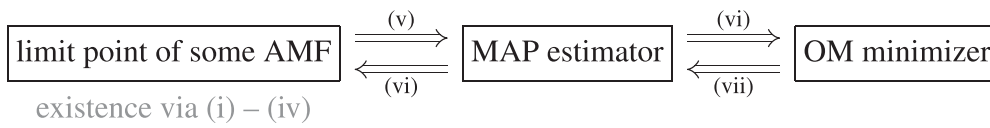
Figure 1. Strategy for proving the existence and equivalence of AMF limit points, MAP estimators and OM minimizers.

that the ζ^{δ_m} do not carry negligible prior ball mass in the asymptotic limit. Second, we show for any sequence $(x_m)_{m\in\mathbb{N}}$ in X that, if either

- $(x_m)_{m\in\mathbb{N}}$ is unbounded or
- $x_m \rightharpoonup \bar{z}$ with $\bar{z} \notin E$ or
- $x_m \rightharpoonup \bar{z} \in E$ but $\|x_m - \bar{z}\|_X \nrightarrow 0$,

then

$$\liminf_{m\to\infty} \mathfrak{R}_\mu^{\delta_m}(x_m, 0) = 0,$$

providing a contradiction for $x_m = \zeta^{\delta_m}$. The three properties described above, as well as (ii), are formulated in condition 2.7 (C1)–(C4) and stated as assumptions in theorem 2.8, which can therefore be seen as a 'shell theorem'. Note that steps (v), (vi) and (vii) then follow in any separable Banach space.

Finally, we prove condition 2.7 (C1)–(C4) and finalize the proof of conjecture 2.3 in the two mentioned cases—section 3 covers the case where X is a Hilbert space (theorem 2.4), while section 4 considers $X = \ell^p$, $1 \leqslant p < \infty$, and diagonal Gaussian measures (theorem 2.5).

Remark 2.6. Apart from providing a 'skeleton' for the proof of conjecture 2.3, the strength of theorem 2.8 lies in its generality: It holds for any separable Banach space and thereby paves the way for future research. Further, remarkably, while condition 2.7 (C1)–(C4) are stated in terms of the prior measure μ, the conclusions are drawn for MAP estimators of μ^y, with assumption 2.1 providing the sufficient conditions for comparability between prior and posterior in order to make this possible.

2.2. A framework for proving existence of MAP estimators

While we use the proof strategy described above to prove theorems 2.4 and 2.5, it paves the way for further research. Note that theorem 2.8 is applicable to any separable Banach space, so this approach can be followed to prove conjecture 2.3 for other classes of Banach spaces.

Condition 2.7. Under assumption 2.1, we introduce the following four conditions:

(C1) *(vanishing condition for unbounded sequences)*—For any null sequence $(\delta_m)_{m\in\mathbb{N}}$ in $\mathbb{R}^+$ and unbounded sequence $(x_m)_{m\in\mathbb{N}}$ in X,

$$\liminf_{m\to\infty} \mathfrak{R}_\mu^{\delta_m}(x_m, 0) = 0.$$

(C2) *(weakly convergent subsequence condition)*—If $(\delta_m)_{m\in\mathbb{N}}$ is a null sequence in $\mathbb{R}^+$ and $(x_m)_{m\in\mathbb{N}}$ is a bounded sequence in X such that there exists $K > 0$ satisfying, for each $m \in \mathbb{N}$, $\mathfrak{R}_\mu^{\delta_m}(x_m, 0) \geqslant K$, then $(x_m)_{m\in\mathbb{N}}$ has a weakly convergent subsequence.

(C3) *(vanishing condition for weak limits outside E)*—For any null sequence $(\delta_m)_{m\in\mathbb{N}}$ in $\mathbb{R}^+$ and weakly convergent sequence $(x_m)_{m\in\mathbb{N}}$ in X with weak limit $\bar{z} \notin E$, $\liminf_{m\to\infty} \mathfrak{R}_\mu^{\delta_m}(x_m, 0) = 0$.[6]

(C4) *(vanishing condition for weakly, but not strongly convergent sequences)*—For any null sequence $(\delta_m)_{m\in\mathbb{N}}$ in $\mathbb{R}^+$ and weakly, but not strongly convergent sequence $(x_m)_{m\in\mathbb{N}}$ in X with weak limit $\bar{z} \in E$, $\liminf_{m\to\infty} \mathfrak{R}_\mu^{\delta_m}(x_m, 0) = 0$.[7]

Theorem 2.8. *Let assumption 2.1 hold and $(\zeta^\delta)_{\delta>0}$ be any asymptotic maximizing family (AMF) in X. Then there exist constants $K > 0$ and $\delta_0 > 0$, such that, for any $0 < \delta < \delta_0$,*

$$\mathfrak{R}_\mu^\delta(\zeta^\delta, 0) \geqslant K. \tag{2.1}$$

It follows that:

(a) *If condition 2.7 (C1)–(C4) hold, $(\zeta^\delta)_{\delta>0}$ is an AMF in X and $(\delta_m)_{m\in\mathbb{N}}$ is a null sequence, then $(\zeta^{\delta_m})_{m\in\mathbb{N}}$ has a subsequence which converges strongly (in X) to an element $\bar{w} \in E$ and any limit point $\bar{z}$ of $(\zeta^\delta)_{\delta>0}$ lies in E and is a MAP estimator for μ^y.*

(b) *If condition 2.7 (C3) holds, then any MAP estimator for μ^y is an element of the Cameron–Martin space E, minimizes the OM functional and is a limit point of some AMF.*

(c) *If condition 2.7 (C3) holds and μ^y has a MAP estimator $\bar{z}$, then any minimizer $\bar{x} \in E$ of the OM functional is also a MAP estimator.*

In particular, if condition 2.7 (C1)–(C4) are satisfied, then conjecture 2.3 holds.

Proof. Due to assumption 2.1(c) and definition 1.7 there exists a family $(\varepsilon^\delta)_{\delta>0}$ such that $\varepsilon^\delta \searrow 0$ for $\delta \searrow 0$, and, for any $0 < \delta \leqslant 1$,

$$\mu^y(B_\delta(\zeta^\delta)) > \frac{1-\varepsilon^\delta}{Z} \cdot \sup_{z\in X}\int_{B_\delta(z)} e^{-\Phi(u)}\mathrm{d}\mu(u) \geqslant \frac{1-\varepsilon^\delta}{Z} \cdot \int_{B_\delta(0)} e^{-\Phi(u)}\mathrm{d}\mu(u)$$

$$\geqslant \frac{1-\varepsilon^\delta}{Z} \cdot \int_{B_\delta(0)} e^{-L(1)}\mathrm{d}\mu(u) = \frac{1-\varepsilon^\delta}{Z} e^{-L(1)}\mu(B_\delta(0)). \tag{2.2}$$

Furthermore, by assumption 2.1(a), for any $z \in X$ and $\delta > 0$,

$$\mu^y(B_\delta(z)) = \frac{1}{Z}\int_{B_\delta(z)} e^{-\Phi(u)}\mathrm{d}\mu(u) \leqslant \frac{e^{-M}}{Z}\mu(B_\delta(z)). \tag{2.3}$$

Choosing $0 < \delta_0 \leqslant 1$ such that $\varepsilon^\delta < 1/2$ for each $0 < \delta < \delta_0$, and denoting $K := e^{M-L(1)}/2$,

$$\mu(B_\delta(\zeta^\delta)) \geqslant Ze^M\mu^y(B_\delta(\zeta^\delta)) \geqslant (1-\varepsilon^\delta)e^{M-L(1)}\mu(B_\delta(0)) \geqslant K\mu(B_\delta(0)),$$

proving (2.1).

Proving (a). Consider the sequence ζ^{δ_m} with $\delta_m \searrow 0$ as $m \to \infty$. Then

(i) condition 2.7 (C1) implies boundedness of $(\zeta^{\delta_m})_{m\in\mathbb{N}}$ in X,

---

[6] This condition corresponds to (Dashti *et al* 2013, lemma 3.7) and (Kretschmann 2019, lemma 4.11). While this is sufficiently strong for our purposes, namely the proofs of the main theorems 2.4 and 2.5, we actually prove the stronger statement with lim sup in place of lim inf both for Hilbert spaces (corollary 3.6) as well as for $X = \ell^p$ (lemma 4.5).

[7] This condition corresponds to (Dashti *et al* 2013, lemma 3.9) and (Kretschmann 2019, lemma 4.13).

(ii) condition 2.7 (C2) implies that $(\zeta^{\delta_m})_{m\in\mathbb{N}}$ has a weakly (in X) convergent subsequence with weak limit point $\bar{w} \in X$.

(iii) condition 2.7 (C3) implies that any weak (in X) limit point $\bar{z} \in X$ of $(\zeta^{\delta_m})_{m\in\mathbb{N}}$ lies in the Cameron–Martin space E.

(iv) condition 2.7 (C4) implies that any weak (in X) limit point $\bar{z} \in E$ of $(\zeta^{\delta_m})_{m\in\mathbb{N}}$ is also a strong (in X) limit point of $(\zeta^{\delta_m})_{m\in\mathbb{N}}$.

In particular, there exists a subsequence of $(\zeta^{\delta_m})_{m\in\mathbb{N}}$ which converges strongly (in X) to some $\bar{w} \in E$. This proves the first part of (a).

Now let $\bar{z}$ be any limit point of $(\zeta^{\delta})_{\delta>0}$ and $(\delta_m)_{m\in\mathbb{N}}$ be such that $(\zeta^{\delta_m})_{m\in\mathbb{N}}$ converges (strongly) to $\bar{z}$. Note that $\bar{z} \in E$ by (iii). We set

$$S := \max\{\|\bar{z}\|_X, \sup_{m\in\mathbb{N}} \|\zeta^{\delta_m}\|_X\}.$$

Using the local Lipschitz constant $L(r)$ for Φ on $B_r(0)$ (see assumption 2.1(c)), we obtain, for any $m \in \mathbb{N}$,

$$\mathfrak{R}_{\mu^y}^{\delta_m}(\zeta^{\delta_m},\bar{z}) = \exp(\Phi(\bar{z})-\Phi(\zeta^{\delta_m})) \frac{\int_{B_{\delta_m}(\zeta^{\delta_m})} e^{\Phi(\zeta^{\delta_m})-\Phi(u)}\,\mathrm{d}\mu(u)}{\int_{B_{\delta_m}(\bar{z})} e^{\Phi(\bar{z})-\Phi(u)}\,\mathrm{d}\mu(u)}$$

$$\leqslant \exp\left(L(S)\cdot\|\zeta^{\delta_m}-\bar{z}\|_X + L(S+\delta_m)\cdot\delta_m + L(S+\delta_m)\cdot\delta_m\right)\mathfrak{R}_{\mu}^{\delta_m}(\zeta^{\delta_m},\bar{z}).$$

Since $\zeta^{\delta_m} \to \bar{z}$ as $m \to \infty$, lemma A.2 and definition 1.7 of AMFs imply

$$\limsup_{m\to\infty} \mathfrak{R}_{\mu^y}^{\delta_m}(\sup,\bar{z}) = \limsup_{m\to\infty} \mathfrak{R}_{\mu^y}^{\delta_m}(\sup,\zeta^{\delta_m})\mathfrak{R}_{\mu^y}^{\delta_m}(\zeta^{\delta_m},\bar{z})$$

$$\leqslant \limsup_{m\to\infty} (1-\varepsilon^{\delta_m})^{-1} \exp\left(L(S)\|\zeta^{\delta_m}-\bar{z}\|_X + 2L(S+\delta_m)\cdot\delta_m\right)\mathfrak{R}_{\mu}^{\delta_m}(\zeta^{\delta_m},\bar{z})$$

$$\leqslant 1. \tag{2.4}$$

If we can show that $\limsup_{\delta\searrow 0}\mathfrak{R}_{\mu^y}^{\delta}(\sup,\bar{z}) \leqslant 1$ (i.e. for any null sequence, not just for $(\delta_m)_{m\in\mathbb{N}}$), then, since $\mathfrak{R}_{\mu^y}^{\delta}(\sup,\bar{z}) \geqslant 1$ for each $\delta > 0$, this implies that in fact $\lim_{\delta\searrow 0}\mathfrak{R}_{\mu^y}^{\delta}(\sup,\bar{z}) = 1$, proving that $\bar{z}$ is a MAP estimator and finalizing the proof. For this purpose assume otherwise, i.e. there exists a null sequence $(\varepsilon_m)_{m\in\mathbb{N}}$ such that $\limsup_{m\to\infty}\mathfrak{R}_{\mu^y}^{\varepsilon_m}(\sup,\bar{z}) > 1$.

With the same argumentation as in (i)–(iv), there exists a subsequence of $(\zeta^{\varepsilon_m})_{m\in\mathbb{N}}$, which, for simplicity, we denote by the same symbol, that converges strongly to some element $\bar{x} \in E$. Similarly to (2.4) we obtain

$$\limsup_{m\to\infty} \mathfrak{R}_{\mu^y}^{\varepsilon_m}(\sup,\bar{x}) \leqslant 1. \tag{2.5}$$

Now, since $\bar{x},\bar{z} \in E$, the property of the OM functional, theorem 1.5, guarantees the existence of the limit $\lim_{\delta\searrow 0}\mathfrak{R}_{\mu^y}^{\delta}(\bar{x},\bar{z})$ and therefore (2.4) implies

$$\lim_{m\to\infty} \mathfrak{R}_{\mu^y}^{\varepsilon_m}(\bar{x},\bar{z}) = \lim_{\delta\searrow 0}\mathfrak{R}_{\mu^y}^{\delta}(\bar{x},\bar{z}) = \lim_{m\to\infty}\mathfrak{R}_{\mu^y}^{\delta_m}(\bar{x},\bar{z}) \leqslant \limsup_{m\to\infty}\mathfrak{R}_{\mu^y}^{\delta_m}(\sup,\bar{z}) \leqslant 1. \tag{2.6}$$

It follows from (2.5) and (2.6) that

$$1 < \limsup_{m\to\infty} \mathfrak{R}_{\mu^y}^{\varepsilon_m}(\sup,\bar{z}) = \limsup_{m\to\infty}\mathfrak{R}_{\mu^y}^{\varepsilon_m}(\sup,\bar{x}) \lim_{m\to\infty}\mathfrak{R}_{\mu^y}^{\varepsilon_m}(\bar{x},\bar{z}) \leqslant 1,$$

which is a contradiction, finalizing the proof.

Proving (b). Now let $\bar{z} \in X$ be any MAP estimator (not necessarily the one obtained as the limit of ζ^{δ_m}). Assuming $\bar{z} \notin E$ and considering the constant sequence $(\bar{z})_{m \in \mathbb{N}}$ (clearly converging to $\bar{z}$), the vanishing condition for weak limits outside E, condition 2.7 (C3), implies that

$$\liminf_{m \to \infty} \mathfrak{R}_\mu^{\delta_m}(\bar{z}, 0) = 0$$

for any null sequence $(\delta_m)_{m \in \mathbb{N}}$. Since the constant family $(\bar{z})_{\delta > 0}$ is an AMF for μ^y by lemma 1.8, (2.1) implies

$$\liminf_{\delta \searrow 0} \mathfrak{R}_\mu^\delta(\bar{z}, 0) \geqslant K > 0.$$

This contradiction proves $\bar{z} \in E$. By definition of MAP estimators and theorem 1.5, it follows for any $z^\star \in E$ that

$$1 = \lim_{\delta \searrow 0} \mathfrak{R}_{\mu^y}^\delta(\bar{z}, \sup) \leqslant \lim_{\delta \searrow 0} \mathfrak{R}_{\mu^y}^\delta(\bar{z}, z^\star) = \exp(I(z^\star) - I(\bar{z})).$$

Hence, $I(z^\star) \geqslant I(\bar{z})$ and $\bar{z}$ is a minimizer of the OM functional. Finally, by lemma 1.8, $\bar{z}$ is also a limit point of the constant AMF $(\bar{z})_{m \in \mathbb{N}}$.

Proving (c). By (b), $\bar{z} \in E$ and minimizes the OM functional I, hence $I(\bar{z}) = I(\bar{x})$. It follows from theorem 1.5 that

$$\lim_{\delta \searrow 0} \mathfrak{R}_{\mu^y}^\delta(\bar{x}, \sup) = \lim_{\delta \searrow 0} \mathfrak{R}_{\mu^y}^\delta(\bar{x}, \bar{z}) \cdot \lim_{\delta \searrow 0} \mathfrak{R}_{\mu^y}^\delta(\bar{z}, \sup) = \exp(I(\bar{z}) - I(\bar{x})) \cdot 1 = 1,$$

proving (c).

In summary, we have shown that each AMF (the existence of some AMF follows from lemma 1.8) has a limit point $\bar{z} \in E$, which is a MAP estimator. Furthermore, each limit point of an AMF lies in E and is a MAP estimator. In addition, any MAP estimator minimizes the OM functional and is a limit point of some AMF. Finally, each minimizer of the OM functional is a MAP estimator. Together, this proves conjecture 2.3.

$\square$

2.3. Some comments on the proof of condition 2.7 (C1)–(C4)

The main obstacle in proving theorems 2.4 and 2.5 is the verification of condition 2.7 (C1)–(C4). Let us shortly summarize one of the main ideas, demonstrated on the derivation of the vanishing condition for unbounded sequences (C1) in the finite-dimensional setting $X = \mathbb{R}^k$, $k \in \mathbb{N}$: our aim is to show that, for any $\delta > 0$ the ratio $\mathfrak{R}_\mu^\delta(x, 0)$ decays to zero as $\|x\|_X \to \infty$. For this purpose we extract a certain prefactor from the integrals in the following way:

$$\mathfrak{R}_\mu^\delta(x, 0) = \frac{\int_{B_\delta(x)} \exp\left(-\frac{1}{2}|u|_E^2\right) \mathrm{d}u}{\int_{B_\delta(0)} \exp\left(-\frac{1}{2}|u|_E^2\right) \mathrm{d}u}$$

$$\leqslant \frac{\sup_{v \in B_\delta(x)} \exp\left(-\frac{1}{2}L(v)\right)}{\inf_{v \in B_\delta(0)} \exp\left(-\frac{1}{2}L(v)\right)} \frac{\int_{B_\delta(x)} \exp\left(-\frac{1}{2}(|u|_E^2 - L(u))\right) \mathrm{d}u}{\int_{B_\delta(0)} \exp\left(-\frac{1}{2}(|u|_E^2 - L(u))\right) \mathrm{d}u}.$$

If L satisfies the following conditions,

(i) there exists $\alpha > 0$ and $\kappa_1, \kappa_2 \geqslant 0$ such that, for each $v \in \mathbb{R}^k$,
   $\|v\|_X^\alpha - \kappa_1 \leqslant L(v) \leqslant \|v\|_X^\alpha + \kappa_2$,
(ii) $|\cdot|_E^2 - L$ is non-negative and convex,

then (ii) implies that, by Anderson's inequality, we can bound the remaining ratio of integrals from above by 1, while (i) implies that, for any fixed $\delta > 0$, the first fraction vanishes as $\|x\|_X \to \infty$.

In separable Hilbert spaces $X = \mathcal{H}$ a function L satisfying (i)–(ii) is not hard to find (in both finite and infinite dimensions) since both $\|\cdot\|_{\mathcal{H}}$ and $|\cdot|_E$ are quadratic. In general separable Banach spaces the large discrepancy between the geometries induced by the norms $\|\cdot\|_X$ and $|\cdot|_E$ strongly complicates the search for such a function L, where convexity is particularly hard to ensure. For $X = \ell^p$, the technical proposition 4.6 guarantees the existence of such a function L. This result together with proposition 4.8 can be seen as the crux to the results presented in this paper.

3. The Hilbert space case: proof of theorem 2.4

In this section we treat the case where $X = \mathcal{H}$ is a Hilbert space, i.e. we prove theorem 2.4. These results have already been presented by Dashti *et al* (2013), with some corrections by Kretschmann (2019). However, both of these manuscripts did not prove the existence of the central object in their proofs, namely the δ-ball maximizing centers $z^\delta = \mathrm{argmax}_x \mu^y(B_\delta(x))$, which seems to be a highly nontrivial issue, see Lambley and Sullivan (2022). This section closes this theoretical gap by working with AMFs ζ^δ defined by definition 1.7 and serves two further purposes:

First, the Hilbert space case provides insight into the main ideas of the proof of conjecture 2.3 with fewer technicalities than the more general case $X = \ell^p$. Second, we use a helpful statement from (Da Prato and Zabczyk 2002), restated in proposition 3.2 below, which simplifies the proofs considerably in comparison to (Dashti *et al* 2013, Kretschmann 2019) and renders the proofs more streamlined.

Notation 3.1. Let $\mathcal{H}$ be an infinite-dimensional separable Hilbert space and $\mu = \mathcal{N}(0, Q)$ a centered and non-degenerate Gaussian measure on $\mathcal{H}$. As the covariance operator Q of μ is a self-adjoint, positive, trace-class operator (Baker 1973), there exists an orthonormal eigenbasis $(e_k)_{k \in \mathbb{N}}$ of Q in which $\mu = \otimes_{k \in \mathbb{N}} \mathcal{N}(0, \sigma_k^2)$ is a product measure of one-dimensional Gaussian measures, where $Q e_k = \sigma_k^2 e_k$ and $\sigma_k > 0$ for each $k \in \mathbb{N}$ and $\sum_{k \in \mathbb{N}} \sigma_k^2 < \infty$. We assume the eigenvalues to be decreasing, i.e. $\sigma_1 \geqslant \sigma_2 \geqslant \ldots$. We write $D = \mathrm{diag}(d_1, d_2, \ldots) := \sum_{k \in \mathbb{N}} d_k e_k \otimes e_k$ for any operator that is diagonal in the basis $(e_k)_{k \in \mathbb{N}}$. Denoting $a_k := \sigma_k^{-2}$ for $k \in \mathbb{N}$, the Cameron–Martin space of μ is given by

$$E = \{z \in \mathcal{H} : \ |z|_E < \infty\}, \qquad |z|_E^2 = \sum_{k=1}^\infty a_k \langle z, e_k \rangle_{\mathcal{H}}^2, \tag{3.1}$$

see Da Prato and Zabczyk (2014, theorem 2.23). Finally, we define the orthogonal projection operators $\Pi^k, \Pi_k \colon \mathcal{H} \to \mathcal{H}, k \in \mathbb{N} \cup \{0\}$, by

$$\Pi^k(x) := \sum_{j=1}^k \langle x, e_j \rangle_{\mathcal{H}} e_j, \qquad \Pi_k(x) := x - \Pi^k(x).$$

Note that $\Pi^0 = 0$ and $\Pi_0 = \mathrm{Id}$.

We start by reciting the following result which will allow us to 'extract an exponential rate' by integrating over a slightly wider Gaussian measure:

Proposition 3.2 (Da Prato and Zabczyk 2002, proposition 1.3.11). *If* $\Gamma\colon \mathscr{H} \to \mathscr{H}$ *is self-adjoint and such that* $Q^{1/2}\Gamma Q^{1/2}$ *is trace class on* $\mathscr{H}$ *and additionally* $\langle x, Q^{1/2}\Gamma Q^{1/2}x\rangle_{\mathscr{H}} < \|x\|^2_{\mathscr{H}}$ *for all* $x \in \mathscr{H}$. *Then for* $\mu = \mathcal{N}(0,Q)$ *and* $\nu = \mathcal{N}(0,(Q^{-1}-\Gamma)^{-1})$ *we have*

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(u) = \frac{\exp\left(-\frac{1}{2}\langle\Gamma u,u\rangle_{\mathscr{H}}\right)}{\sqrt{\det(I-Q^{1/2}\Gamma Q^{1/2})}}.$$

Remark 3.3. In one dimension this boils down to the following: let $\sigma > 0$ and $\mu = \mathcal{N}(0,\sigma^2)$. Then, for any $\gamma < \sigma^{-2}$,

$$\mu(A) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_A \exp\left(-\gamma^2\frac{x^2}{2}\right)\exp\left(-\frac{x^2}{2\left(\frac{\sigma^2}{1-\gamma^2\sigma^2}\right)}\right)\mathrm{d}x$$

$$= \int_A \frac{\exp\left(-\gamma^2\frac{x^2}{2}\right)}{\sqrt{1-\gamma^2\sigma^2}}\mathrm{d}\nu(x),$$

where $\nu = \mathcal{N}(0,\frac{\sigma^2}{1-\gamma^2\sigma^2}) = \mathcal{N}(0,(\sigma^{-2}-\gamma^2)^{-1})$.

Then we can re-prove the following lemma (as already stated in (Dashti *et al* 2013) and (Kretschmann 2019)):

Lemma 3.4 (Dashti *et al* 2013, lemma 3.6). *Let assumption 2.1 hold and* $X = \mathscr{H}$ *be a separable Hilbert space. Then, using notation 3.1, for any* $\delta > 0$ *and* $z \in \mathscr{H}$, *and* $n \in \mathbb{N}$,

$$\mathfrak{R}^{\delta}_{\mu}(z,0) \leqslant \exp\left(-\frac{a_n}{2}\left[(\|\Pi_{n-1}z\|_{\mathscr{H}}-\delta)^2-\delta^2\right]\right).$$

Proof. Using notation 3.1, for arbitrary $n \geqslant n_0$, let $\Gamma = \mathrm{diag}(0,\ldots,0,r,\ldots,)$ with entries $0 < r < a_n$ starting at position n, such that $Q^{-1}-\Gamma = \mathrm{diag}(a_1,\ldots,a_{n-1},a_n-r,a_{n+1}-r,\ldots)$ is a valid precision (i.e. inverse covariance) operator of a Gaussian measure on $\mathscr{H}$. This means that $\langle x,\Gamma x\rangle = r\|\Pi_{n-1}x\|^2_X$. This choice of Γ fulfills the conditions of proposition 3.2: first, $(Q^{-1}-\Gamma)^{-1}$ is a valid covariance operator:

$$\sum_{i=n}^{\infty}(a_i-r)^{-1} = \sum_{i=n}^{\infty}\frac{a_i^{-1}}{1-ra_i^{-1}} \leqslant \frac{1}{1-ra_n^{-1}}\sum_{i=n}^{\infty}a_i^{-1} = \frac{a_n}{a_n-r}\sum_{i=n}^{\infty}\sigma_i^2 < \infty.$$

Second, since Q is trace class (Baker 1973), so is

$$Q^{1/2}\Gamma Q^{1/2} = \mathrm{diag}(0,\ldots,0,r\sigma_n^2,r\sigma_{n+1}^2,\ldots).$$

Finally, as $r < a_n = \sigma_n^{-2}$, and $\sigma_m^2 \leqslant \sigma_n^2$ for $m > n$, we also have that $r\sigma_m^2 \leqslant 1$ for all $m \geqslant n$, hence $\langle x,Q^{1/2}\Gamma Q^{1/2}x\rangle \leqslant \|x\|^2_X$.

Thus, with $\nu = \mathcal{N}(0,(Q^{-1}-\Gamma)^{-1})$, proposition 3.2 implies for any $\delta > 0$:

$$\mathfrak{R}^{\delta}_{\mu}(z,0) = \frac{\int_{B_{\delta}(z)}e^{-\frac{1}{2}\langle x,\Gamma x\rangle_{\mathscr{H}}}\mathrm{d}\nu(x)}{\int_{B_{\delta}(0)}e^{-\frac{1}{2}\langle x,\Gamma x\rangle_{\mathscr{H}}}\mathrm{d}\nu(x)}$$

$$\leqslant \frac{\exp(-\frac{r}{2}(\|\Pi_{n-1}z\|_{\mathscr{H}}-\delta)^2)}{\exp(-\frac{r}{2}\delta^2)}\frac{\int_{B_{\delta}(z)}\mathrm{d}\nu(x)}{\int_{B_{\delta}(0)}\mathrm{d}\nu(x)}$$

$$\leqslant \exp\left(-\frac{r}{2}\left[(\|\Pi_{n-1}z\|_{\mathscr{H}}-\delta)^2-\delta^2\right]\right)$$

due to Anderson's inequality (theorem A.4 with $\gamma = \nu$, $A = B_\delta(0)$ and $a = z$). Since above inequality holds for any $0 < r < a_n$, it also holds for $r = a_n$ by continuity, and the claim follows. ☐

Corollary 3.5. *Let assumption 2.1 hold and $X = \mathscr{H}$ be a separable Hilbert space. Then the vanishing condition for unbounded sequences, condition 2.7 (C1), holds.*

Proof. Let $(\delta_m)_{m \in \mathbb{N}}$ be a null sequence in $\mathbb{R}^+$ and $(x_m)_{m \in \mathbb{N}}$ be an unbounded sequence in X. We have to prove that for any $\varepsilon > 0$ and any $m \in \mathbb{N}$ there exists a $m^\star \geqslant m$ such that

$$\mathfrak{R}_\mu^{\delta_{m^\star}(x_{m^\star}, 0)} \leqslant \varepsilon.$$

Indeed, for arbitrary $\varepsilon > 0$ and $m \in \mathbb{N}$ there exists $M > 0$ such that $\frac{a_1 M^2}{4} \geqslant \log \varepsilon^{-1}$. Since (δ_m) is a null sequence, there exists $m_1 \geqslant m$ such that for all $n \geqslant m_1$, $\delta_n < M/4$. By unboundedness of $(x_m)_m$ we can find a $m^\star \geqslant m_1 \geqslant m$ such that $\|x_{m^\star}\|_{\mathscr{H}} \geqslant M$. Then, by lemma 3.4,

$$\begin{aligned}
\mathfrak{R}_\mu^{\delta_{m^\star}(x_{m^\star}, 0)} &\leqslant \exp\left(-\frac{a_1}{2} \left[(\|x_{m^\star}\|_{\mathscr{H}} - \delta_{m^\star})^2 - \delta_{m^\star}^2 \right] \right) \\
&\leqslant \exp\left(-\frac{a_1}{2} \left[\frac{9M^2}{16} - \frac{M^2}{16} \right] \right) = \exp\left(-\frac{a_1 M^2}{4} \right) \leqslant \varepsilon.
\end{aligned}$$

☐

Similarly we can shorten the proof of the following lemma:

Corollary 3.6 (Dashti *et al* (2013, lemma 3.7), Kretschmann (2019, lemma 4.11)). *Let assumption 2.1 hold and $X = \mathscr{H}$ be a separable Hilbert space. Then the vanishing condition for weak limits outside E, condition 2.7 (C3), is satisfied.*

Proof. We use notation 3.1 throughout the proof.

Let $(\delta_m)_{m \in \mathbb{N}}$ be a null sequence in $\mathbb{R}^+$ and $(x_m)_{m \in \mathbb{N}}$ be a weakly convergent sequence with weak limit $\bar{z} \notin E$.

$$\mathfrak{R}_\mu^{\delta_m(x_m, 0)} \leqslant \varepsilon.$$

Let $\varepsilon > 0$. Since $\bar{z} \notin E$, $|\Pi^n \bar{z}|_E \to \infty$ as $n \to \infty$ by (3.1), hence there exists $n \in \mathbb{N}$ (which we fix from now on) such that

$$|\Pi^n \bar{z}|_E \geqslant 4, \qquad \exp\left(-\frac{3}{64} |\Pi^n \bar{z}|_E^2 \right) < \varepsilon. \tag{3.2}$$

Note that $\Gamma := \mathrm{diag}(a_1/2, a_2/2, \ldots, a_n/2, 0, 0, \ldots)$ is a valid choice for the operator Γ in proposition 3.2 and observe that

$$\langle x, \Gamma x \rangle_{\mathscr{H}} = \frac{1}{2} |\Pi^n x|_E^2, \qquad x \in \mathscr{H}. \tag{3.3}$$

Since weak convergence $x_m \rightharpoonup \bar{z}$ implies componentwise convergence, there exists $m_1 \in \mathbb{N}$ such that, for any $m \geqslant m_1$, $|\Pi^n (\bar{z} - x_m)|_E \leqslant 1$. Since $(\delta_m)_{m \in \mathbb{N}}$ is a null sequence, there exists $m^\star \geqslant m_1$ such that, for each $m \geqslant m^\star$, $\delta_m^2 \leqslant \sigma_n^2/n$. It follows from (3.2) for any $m \geqslant m^\star$, any $z \in B_{\delta_m}(x_m)$ and any $w \in B_{\delta_m}(0)$, denoting $x_{m,j}, z_j, w_j$ for the jth component of x_m, z, w, that

(i) $|\Pi^n (x_m - z)|_E^2 = \sum_{j=1}^n \sigma_j^{-2} |x_{m,j} - z_j|^2 \leqslant \sum_{j=1}^n \sigma_n^{-2} \delta_m^2 \leqslant \sum_{j=1}^n n^{-1} = 1$;

(ii) $|\Pi^n z|_E \geqslant \frac{1}{2} |\Pi^n \bar{z}|_E + \underbrace{\frac{1}{2} |\Pi^n \bar{z}|_E}_{\geqslant 2} - \underbrace{|\Pi^n (\bar{z} - x_m)|_E}_{\leqslant 1} - \underbrace{|\Pi^n (x_m - z)|_E}_{\leqslant 1} \geqslant \frac{1}{2} |\Pi^n \bar{z}|_E$;

(iii) $|\Pi^n w|_E^2 = \sum_{j=1}^n \sigma_j^{-2} |w_j|^2 \leqslant \sum_{j=1}^n \sigma_n^{-2} \delta_m^2 \leqslant \sum_{j=1}^n n^{-1} = 1 \leqslant \frac{1}{16} |\Pi^n \bar{z}|_E^2$.

Using (3.3) and Anderson's inequality (theorem A.4) applied to the Gaussian measure ν on $\mathscr{H}$ as defined in proposition 3.2, this implies, for any $m \geqslant m^\star$,

$$
\begin{aligned}
\mathfrak{R}_\mu^{\delta_m}(x_m, 0) &= \frac{\int_{B_{\delta_m}(x_m)} \exp\left(-\frac{1}{2}\langle z, \Gamma z\rangle_{\mathscr{H}}\right) \mathrm{d}\nu(z)}{\int_{B_{\delta_m}(0)} \exp\left(-\frac{1}{2}\langle w, \Gamma w\rangle_{\mathscr{H}}\right) \mathrm{d}\nu(w)} \\
&\leqslant \exp\left(\frac{1}{4} \sup_{w \in B_{\delta_m}(0)} |\Pi^n w|_E^2 - \frac{1}{4} \inf_{z \in B_{\delta_m}(x_m)} |\Pi^n z|_E^2\right) \frac{\nu(B_{\delta_m}(x_m))}{\nu(B_{\delta_m}(0))} \\
&\leqslant \exp\left(\frac{1}{64} |\Pi^n \bar{z}|_E^2 - \frac{1}{16} |\Pi^n \bar{z}|_E^2\right) \\
&= \exp\left(-\frac{3}{64} |\Pi^n \bar{z}|_E^2\right) < \varepsilon,
\end{aligned}
$$

proving the claim. $\qquad\square$

Corollary 3.7 (Dashti *et al* 2013, lemma 3.9 and Kretschmann 2019, lemma 4.13). *Let assumption 2.1 hold and $X = \mathscr{H}$ be a separable Hilbert space. Then the vanishing condition for weakly, but not strongly convergent sequences, condition 2.7 (C4), is satisfied.*

Proof. We use notation 3.1 throughout the proof. Let $(\delta_m)_{m \in \mathbb{N}}$ be a null sequence in $\mathbb{R}^+$ and $(x_m)_{m \in \mathbb{N}}$ converge weakly, but not strongly to $\bar{z} \in E$. We will show that, for any $\varepsilon > 0$ and $m_1 \in \mathbb{N}$, there exists $m^\star \geqslant m_1$ such

$$
\mathfrak{R}_\mu^{\delta_{m^\star}(x_{m^\star}, 0)} \leqslant \varepsilon.
$$

Now let $\varepsilon > 0$ and $m_1 \in \mathbb{N}$. Since weak convergence $x_m \rightharpoonup \bar{z}$ implies $\|\bar{z}\|_{\mathscr{H}} \leqslant \liminf_{m \to \infty} \|x_m\|_{\mathscr{H}}$ and as the convergence is not strong by assumption, the Radon–Riesz property guarantees the existence of $c > 0$ such that

$$
\limsup_{m \to \infty} \|x_m\| > \|\bar{z}\|_{\mathscr{H}} + c. \tag{3.4}
$$

(Otherwise, $\lim_{m \to \infty} \|x_m\| = \|\bar{z}\|_{\mathscr{H}}$, in which case weak convergence implies strong convergence.) Since $a_k \to \infty$ as $k \to \infty$, there exists $n \in \mathbb{N}$ (which we fix from now on) such that $a_n \geqslant -24c^{-2} \log \varepsilon$. Since $(\delta_m)_{m \in \mathbb{N}}$ is a null sequence and weak convergence $x_m \rightharpoonup \bar{z}$ implies componentwise convergence, (3.4) guarantees the existence of $m^\star \geqslant m_1$ such that $\delta_{m^\star} \leqslant c/6$, $\|\Pi^n(\bar{z} - x_{m^\star})\|_{\mathscr{H}} < c/2$ and $\|x_{m^\star}\|_{\mathscr{H}} > \|\bar{z}\|_{\mathscr{H}} + c$. This implies

$$
\|\Pi_n x_m\|_{\mathscr{H}} \geqslant \|x_m\|_{\mathscr{H}} - \|\Pi^n x_m\|_{\mathscr{H}} > \|\bar{z}\|_{\mathscr{H}} + c - \|\Pi^n(x_m - \bar{z})\|_{\mathscr{H}} - \|\bar{z}\|_{\mathscr{H}} \geqslant c/2
$$

and lemma 3.4 yields

$$
\mathfrak{R}_\mu^{\delta_{m^\star}(x_{m^\star}, 0)} \leqslant \exp\left(-\frac{a_n}{2}\left[(\|\Pi_n x_{m^\star}\|_{\mathscr{H}} - \delta_{m^\star})^2 - \delta_{m^\star}^2\right]\right) \leqslant \exp\left(-\frac{a_n c^2}{24}\right) \leqslant \varepsilon.
$$

$\qquad\square$

Proof of theorem 2.4. By lemma 3.4, corollaries 3.6 and 3.7, condition 2.7 (C1), (C3) and (C4) are fulfilled, while the weakly convergent subsequence condition (C2) follows from the reflexivity of $\mathscr{H}$. Hence, all statements follow from theorem 2.8. $\qquad\square$

4. The case $X = \ell^p$: proof of theorem 2.5

In this section we will extend the results in section 3 to the spaces $X = \ell^p$, $1 \leqslant p < \infty$, i.e. we will prove theorem 2.5. Note that theorem 2.5 is an actual generalization of theorem 2.4 since the covariance structure in a Hilbert space can always be 'diagonalized' by choosing an orthonormal eigenbasis of the covariance operator, which is a consequence of the Karhunen–Loève expansion (Sprungk 2017, theorem 2.21). In other words, the Hilbert space case $(\mathscr{H}, \mu)$ with an arbitrary non-degenerate Gaussian measure μ is equivalent to the case $(\ell^2, \otimes \mathscr{N}(0, \sigma_k^2))$, where σ_k^2 are the corresponding eigenvalues (note that the Cameron–Martin space E respects this equivalence due to (3.1)), and the setting considered in this manuscript corresponds to the canonical generalization from ℓ^2 to ℓ^p, $1 \leqslant p < \infty$.

While our proof strategy is quite similar to the one in (Dashti *et al* 2013), the strong discrepancy between the geometries of the unit balls in E and $X = \ell^p$ for $p \neq 2$ poses a strong obstacle when attempting to extract an exponential decay rate out of the ratio $\mathfrak{R}_\mu^\delta(z, 0)$ with fixed $\delta > 0$, similar to the statement of lemma 3.4 in the Hilbert space case.

To see exactly why this is problematic, let us reiterate on the crucial line in the proof of lemma 3.4. We set $n = 1$ for simplicity, and we focus on the finite-dimensional case (or finite-dimensional approximation to the infinite-dimensional case) which allows to write the integrals with respect to Lebesgue measure. Due to the fact that the Hilbert space norm coincides with an (unweighted) ℓ^2-norm, we can extract a multiple of the Hilbert space norm out of the integral, where $\delta > 0$, $z \in \mathscr{H}$ and $r > 0$ is a sufficiently small constant:

$$
\begin{aligned}
\mathfrak{R}_\mu^\delta(z, 0) &= \frac{\int_{B_\delta(z)} \exp\left(-\frac{1}{2}\left[a_1 x_1^2 + \cdots + a_N x_N^2\right]\right) \mathrm{d}x}{\int_{B_\delta(0)} \exp\left(-\frac{1}{2}\left[a_1 x_1^2 + \cdots + a_N x_N^2\right]\right) \mathrm{d}x} \\
&\leqslant \frac{\sup_{x \in B_\delta(z)} \exp\left(-\frac{r}{2}\|x\|_{\mathscr{H}}^2\right)}{\inf_{x \in B_\delta(0)} \exp\left(-\frac{r}{2}\|x\|_{\mathscr{H}}^2\right)} \\
&\quad \cdot \frac{\int_{B_\delta(z)} \exp\left(-\frac{1}{2}\left[(a_1 - r)x_1^2 + \cdots + (a_N - r)x_N^2\right]\right) \mathrm{d}x}{\int_{B_\delta(0)} \exp\left(-\frac{1}{2}\left[(a_1 - r)x_1^2 + \cdots + (a_N - r)x_N^2\right]\right) \mathrm{d}x} \\
&\leqslant \frac{\exp\left(-\frac{r}{2}(\|z\|_{\mathscr{H}} - \delta)^2\right)}{\exp\left(-\frac{r}{2}\delta^2\right)}
\end{aligned}
$$

where the second factor (the ratio of the remaining integrals) can be bounded by 1 due to Anderson's inequality (theorem A.3) under some prerequisites: first, the ambient space norm $\|\cdot\|_{\mathscr{H}}$ needs to be dominated by (a multiple of) the Cameron–Martin norm such that the integrand is integrable—this is also true for the Banach space case, simply by compact embedding of E in X. Second, the function $|\cdot|_E - r\|\cdot\|_{\mathscr{H}}$ needs to be convex. This is trivially the case in the Hilbert space case due to this difference being a positive definite quadratic, but does not generalize to the Banach space case. Indeed, $|\cdot|_E - \beta\|\cdot\|_p$ is not convex for $p = 1$ and any $\beta > 0$. This issue is solved (in the general ℓ^p case) by proposition 4.6, which demonstrates how to find functions L such that $|\cdot|_E^2 - \beta L(\cdot)$ is convex and L is a suitable surrogate of the ambient space norm $\|\cdot\|_p$, see figure 2 for an illustration.

Proposition 4.8 then leverages this result towards a generalization of lemma 3.4 in the ℓ^p case, after which the proof of validity of condition 2.7 and subsequently theorem 2.5 is more or less straight-forward.

When working in sequence spaces $X \subseteq \mathbb{R}^{\mathbb{N}}$, such as ℓ^p spaces, one important technique (Dashti *et al* 2013, Agapiou *et al* 2018, Ayanbayev *et al* 2021b) is to consider
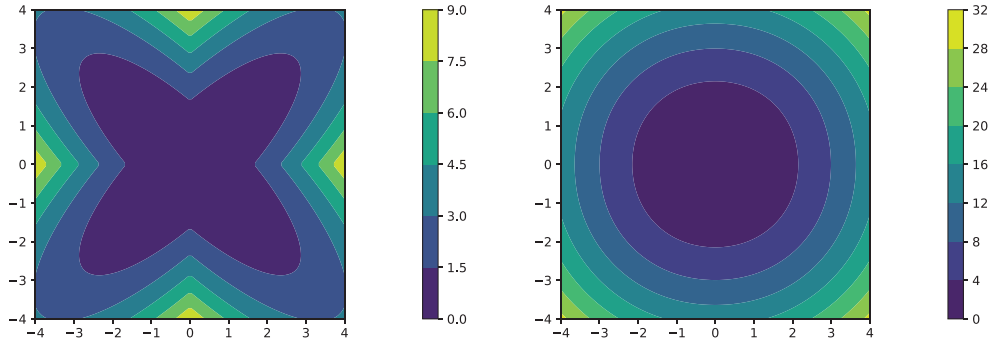
Figure 2. Visualization of the 2d case, $X = \ell^1$ and $\mu = \mathcal{N}(0,1) \otimes \mathcal{N}(0,1)$. *Left:* Plot of the function $(x_1, x_2) \mapsto x_1^2 + x_2^2 - \beta(|x_1| + |x_2|)^2$ for a specific $\beta > 0$. The level sets show that this function is non-convex (this is indeed true for any $\beta > 0$). *Right:* Plot of the function $(x_1, x_2) \mapsto x_1^2 + x_2^2 - \beta L(x_1, x_2)$ for suitable β, which is seen to be convex.

finite-dimensional approximations of $\mu(B_\delta(x))$, $x \in X$. For this purpose, we introduce the following notation:

Assumption 4.1. *We consider $X = \ell^p := \ell^p(\mathbb{N})$ with $1 \leqslant p < \infty$ together with $\mu = \otimes_{j \in \mathbb{N}} \mathcal{N}(0, \sigma_j^2)$, a non-degenerate centered Gaussian measure on X with diagonal covariance structure, where $\sigma_1 \geqslant \sigma_2 \geqslant \ldots > 0$ and $\sum_{j \in \mathbb{N}} \sigma_j^p < \infty$.*

Remark 4.2. The condition $\sum_{j \in \mathbb{N}} \sigma_j^p < \infty$ is a necessary condition for $\mu(X) = 1$ (i.e. samples $(x_i)_{i \in \mathbb{N}}$ are almost surely in ℓ^p), see (Ayanbayev *et al* 2021b, lemma B.3).

Notation 4.3. Let assumption 4.1 hold. Define

$$\alpha := \min(p, 2), \qquad q := \max(p, 2(p-1)^2), \qquad S := \left(\sum_{j \in \mathbb{N}} \sigma_j^p \right)^{1/p}.$$

Further, for $k, K \in \mathbb{N} \cup \{0\}$ with $K > k$ define the projection operators $P^k \colon \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^k$, $P_k \colon \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$, $P_k^K \colon \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{K-k}$ and $P^{-k} \colon \mathbb{R}^k \to \mathbb{R}^{\mathbb{N}}$ by

$$P^k(x) := (x_1, \ldots, x_k), \qquad P_k(x) := (x_{k+1}, x_{k+2}, \ldots),$$
$$P_k^K(x) := (x_{k+1}, \ldots, x_K), \qquad P^{-k}(u) := (u_1, \ldots, u_k, 0, 0, \ldots),$$

where $P^k := 0$ for $k = 0$. Accordingly, we define, for any $u \in \mathbb{R}^k$ and $v \in \mathbb{R}^K$,

- $|u|_{E^k} := \sum_{j=1}^k \sigma_j^{-2} u_j^2, \qquad |v|_{E_k^K} := \sum_{j=k+1}^K \sigma_j^{-2} v_j^2,$
- $B_\delta^k(u) := \{ w \in \mathbb{R}^k \mid \|w - u\|_p < \delta \},$
- $\mu_k = \otimes_{j=1}^k \mathcal{N}(0, \sigma_j^2).$

Note that $\frac{1}{2} | \cdot |_{E^k}$ is the negative log density of μ_k.

Lemma 4.4. *If assumption 4.1 holds, then the Cameron–Martin space of (ℓ^p, μ) is given by $E = \{ z \in \ell^p : |z|_E < \infty \}$ where $|z|_E^2 := \sum_{k=1}^\infty \frac{z_k^2}{\sigma_k^2}.$*

Proof. By (Bogachev 1998, lemma 3.2.2), we may consider μ as a Gaussian measure on a Hilbert space $\mathscr{H} \supseteq X$, into which X is continuously and linearly embedded, without changing

17

the Cameron–Martin space or its norm. If $p \leqslant 2$, X is continuously embedded in $\mathcal{H} = \ell^2 \supset X$, since $\|\cdot\|_2 \leqslant \|\cdot\|_p$. For $p > 2$, this can be accomplished by choosing any positive sequence $b \in \ell^{\frac{p}{p-2}}$ and $\mathcal{H} := \{x \in \mathbb{R}^{\mathbb{N}} : \|x\|_{\mathcal{H}}^2 := \sum_{k \in \mathbb{N}} b_k x_k^2 < \infty\}$, since, by Hölder's inequality,

$$\|x\|_{\mathcal{H}}^2 = \sum_{k \in \mathbb{N}} b_k x_k^2 \leqslant \|b\|_{\frac{p}{p-2}} \cdot \|(x_k^2)_{k \in \mathbb{N}}\|_{\frac{p}{2}} \leqslant \|b\|_{\frac{p}{p-2}} \cdot \|x\|_p^2.$$

The Cameron–Martin space and its norm for both X and $\mathcal{H}$ are therefore given by the well-known formulas (3.1), see e.g. (Da Prato and Zabczyk 2014, theorem 2.23), proving the claim.

$\square$

In order to prove theorem 2.5, we will again proceed by showing conditions 2.7 (C1)–(C4) and then applying theorem 2.8. We start by showing the vanishing condition for weak limits outside E (C3), while the vanishing condition for unbounded sequences (C1) and the vanishing condition for weakly, but not strongly convergent sequences (C4) will require some additional work (propositions 4.6 and 4.8).

Lemma 4.5. *Under assumptions 2.1 and 4.1, for any family $(x^\delta)_{0 < \delta < 1}$ in X and for any $\bar{z} \in X \setminus E$, such that $x^\delta \rightharpoonup \bar{z}$ converges weakly as $\delta \searrow 0$, we have*

$$\limsup_{\delta \searrow 0} \mathfrak{R}_\mu^\delta(x^\delta, 0) = 0.$$

In particular, the vanishing condition for weak limits outside E, condition 2.7 (C3), is satisfied.

Proof. We use notation 4.3 throughout the proof. Let $(x^\delta)_{0 < \delta < 1}$ be a family in X and $\bar{z} \in X \setminus E$ such that $x^\delta \rightharpoonup \bar{z}$ converges weakly as $\delta \searrow 0$. Let $0 < \varepsilon < 1$ be arbitrary and $A := \sqrt{8 \log(2/\varepsilon)}$. We proceed in four steps.

Step 1: There exist $K_1 \in \mathbb{N}$ and $\delta_1 > 0$ such that, for each $u \in B_{\delta_1}^{K_1}(P^{K_1}\bar{z})$, $|u|_{E^{K_1}} \geqslant A$.

In order to see this, we assume the contrary, i.e. for each $K_1 \in \mathbb{N}$ and $\delta_1 > 0$, there exists $u \in B_{\delta_1}^{K_1}(P^{K_1}\bar{z})$ with $|u|_{E^{K_1}} < A$. Then, for each $m \in \mathbb{N}$ (choosing $K_1 = m$ and $\delta_1 = m^{-1}$), there exists $u^{(m)} \in B_{m^{-1}}^m(P^m\bar{z})$ with $|u^{(m)}|_{E^m} < A$.

Since $(P^{-m}u^{(m)})_{m \in \mathbb{N}}$ is bounded in E by A, it has a weakly convergent (in E) subsequence, which, for simplicity, we also denote by $(P^{-m}u^{(m)})_{m \in \mathbb{N}}$, with weak limit $\bar{u} \in E$. Further, since $u^{(m)} \in B_{m^{-1}}^m(P^m\bar{z})$ for each $m \in \mathbb{N}$, $P^{-m}u^{(m)} \to \bar{z}$ strongly in X as $m \to \infty$:

$$\|P^{-m}u^{(m)} - \bar{z}\|_p^p = \|u^{(m)} - P^m\bar{z}\|_p^p + \|0 - P_m\bar{z}\|_p^p < m^{-p} + \|P_m\bar{z}\|_p^p \xrightarrow[m \to \infty]{} 0.$$

By considering each component $j \in \mathbb{N}$ separately, weak convergence in E and (strong) convergence in X imply

$$u_j^{(m)} \xrightarrow[m \to \infty]{} \bar{u}_j, \qquad u_j^{(m)} \xrightarrow[m \to \infty]{} \bar{z}_j, \qquad j \in \mathbb{N}.$$

Hence, by the uniqueness of the limit (in $\mathbb{R}$), we obtain the contradiction $E \ni \bar{u} = \bar{z} \notin E$.

Step 2: There exists $0 < \delta_2 < \delta_1/2$ such that, for each $0 < \delta < \delta_2$ and each $u \in B_\delta^{K_1}(P^{K_1}x^\delta)$, we have that $|u|_{E^{K_1}} \geqslant A$.

This can be seen as follows: since $x^\delta \rightharpoonup \bar{z}$ converges weakly (and therefore componentwise) in X, there exists $0 < \delta_2 < \delta_1/2$ such that, for each $0 < \delta < \delta_2$, we have that $\|P^{K_1}x^\delta - P^{K_1}\bar{z}\|_p < \delta_1/2$.

Hence, for each $0 < \delta < \delta_2$ and each $u \in B_\delta^{K_1}(P^{K_1}x^\delta)$,

$$\|u - P^{K_1}\bar{z}\|_p \leqslant \|u - P^{K_1}x^\delta\|_p + \|P^{K_1}x^\delta - P^{K_1}\bar{z}\|_p < \delta + \frac{\delta_1}{2} \leqslant \delta_1,$$

i.e. $B_\delta^{K_1}(P^{K_1}x^\delta) \subseteq B_{\delta_1}^{K_1}(P^{K_1}\bar{z})$ for each $0 < \delta < \delta_2$, and the claim follows from Step 1.

Step 3: There exists $0 < \delta^\star < \delta_2$ such that, for each $\delta < \delta^\star$ and each $u \in B_\delta^{K_1}(0)$, we have $|u|_{E^{K_1}} \leqslant A/\sqrt{2}$.

This is evident from the fact that $|\cdot|_{E^{K_1}}$ and $\|P^{K_1}\cdot\|$ are equivalent norms on the (finite-dimensional) vector space $P^{K_1}X$.

Step 4: For each $0 < \delta < \delta^\star$, $\mathfrak{R}_\mu^\delta(x^\delta, 0) \leqslant \varepsilon$, finalizing the proof.

Let $0 < \delta < \delta^\star$. For any $x \in X$, since $B_\delta(x) = \bigcap_{k\in\mathbb{N}} B_\delta^k(P^k x) \times \mathbb{R}^{\mathbb{N}\setminus\{1,\dots,k\}}$, the continuity of measures implies that $\mu_k(B_\delta^k(P^k x)) \to \mu(B_\delta(x))$. Hence, there exists $k > K_1$ such that

$$\left| \mathfrak{R}_\mu^\delta(x^\delta, 0) - \frac{\mu_k(B_\delta^k(P^k x^\delta))}{\mu_k(B_\delta^k(P^k 0))} \right| \leqslant \frac{\varepsilon}{2}.$$

Since, for any $x \in X$, $\mathbb{R}^k \ni v \in B_\delta^k(P^k x)$ implies $P^{K_1} v \in B_\delta^{K_1}(P^{K_1} x)$, it follows from Steps 2 and 3 that

$$
\begin{aligned}
\mathfrak{R}_\mu^\delta(x^\delta, 0) &\leqslant \left| \mathfrak{R}_\mu^\delta(x^\delta, 0) - \frac{\mu_k(B_\delta^k(P^k x^\delta))}{\mu_k(B_\delta^k(P^k 0))} \right| + \frac{\mu_k(B_\delta^k(P^k x^\delta))}{\mu_k(B_\delta^k(P^k 0))} \\
&\leqslant \frac{\varepsilon}{2} + \frac{\int_{B_\delta^k(P^k x^\delta)} \exp\left(-\frac{1}{2} \sum_{j=1}^k \frac{u_j^2}{\sigma_j^2} \right) du}{\int_{B_\delta^k(0)} \exp\left(-\frac{1}{2} \sum_{j=1}^k \frac{u_j^2}{\sigma_j^2} \right) du} \\
&\leqslant \frac{\varepsilon}{2} + \frac{\sup_{v\in B_\delta^k(P^k x^\delta)} \exp\left(-\frac{1}{4} |P^{K_1} v|_{E^{K_1}}^2 \right)}{\inf_{v\in B_\delta^k(0)} \exp\left(-\frac{1}{4} |P^{K_1} v|_{E^{K_1}}^2 \right)} \\
&\quad \cdot \frac{\int_{B_\delta^k(P^k x^\delta)} \exp\left(-\frac{1}{4} \sum_{j=1}^{K_1} \frac{u_j^2}{\sigma_j^2} - \frac{1}{2} \sum_{j=K_1+1}^k \frac{u_j^2}{\sigma_j^2} \right) du}{\int_{B_\delta^k(0)} \exp\left(-\frac{1}{4} \sum_{j=1}^{K_1} \frac{u_j^2}{\sigma_j^2} - \frac{1}{2} \sum_{j=K_1+1}^k \frac{u_j^2}{\sigma_j^2} \right) du} \\
&\leqslant \frac{\varepsilon}{2} + \exp\left(-\frac{A^2}{4} + \frac{A^2}{8} \right) \cdot 1 = \varepsilon,
\end{aligned}
$$

where we bounded the last ratio of integrals by 1 using Anderson's inequality (theorem A.3).

$\square$

As explained above, the following proposition implements a convexification of the function $|\cdot|_E - \beta\|\cdot\|_p$, which is necessary for the application of Anderson's inequality in the proof of proposition 4.8:

Proposition 4.6. *Using notation 4.3, let $1 \leqslant p < \infty$, let $k \in \mathbb{N}$ and $\rho \in \mathbb{R}^k$ with $\rho_1 \geqslant \dots \geqslant \rho_k > 0$. Further, let $\gamma > 0$, let $\beta_* := \frac{2\gamma^{2-\alpha}}{q\rho_1^\alpha}$ and let $0 \leqslant \beta < \beta_*$. Then the functions $L_{\rho,\gamma}, f_{\rho,\beta,\gamma} \colon \mathbb{R}^k \to \mathbb{R}$ given by*

$$
L_{\rho,\gamma}(x) := \begin{cases} \sum_{j=1}^k (\gamma^2 \rho_j^2 + x_j^2)^{p/2} - (\gamma\rho_j)^p & \text{if } 1 \leqslant p \leqslant 2, \\ \|x\|_p^2 & \text{if } 2 < p < \infty, \end{cases}
$$

$$
f_{\rho,\beta,\gamma}(x) = \sum_{j=1}^k \frac{x_j^2}{\rho_j^2} - \beta L_{\rho,\gamma}(x),
$$

satisfy

(a) $\|x\|_p^\alpha - \gamma^\alpha \|\rho\|_p^\alpha \leqslant L_{\rho,\gamma}(x) \leqslant \|x\|_p^\alpha$ *for any $x \in \mathbb{R}^k$;*
(b) $f_{\rho,\beta,\gamma}$ *is non-negative;*
(c) $f_{\rho,\beta,\gamma}$ *is convex.*

Proof. Recall that, for $0 \leqslant p_1 \leqslant p_2 < \infty$, and $v \in \mathbb{R}^n$, $n \in \mathbb{N}$,

$$\|v\|_{p_1} \geqslant \|v\|_{p_2}. \tag{4.1}$$

While (a) is trivial for $p > 2$, it follows for $1 \leqslant p \leqslant 2$ directly from the inequalities $a^q \leqslant (a+b)^q \leqslant a^q + b^q$ for any $a, b \geqslant 0$ and $q \leqslant 1$, where the second inequality is a consequence of (4.1) for $v = (a, b)$:

$$(a+b)^q = \|(a,b)\|_1^q \leqslant \|(a,b)\|_q^q = a^q + b^q.$$

For (b), note that, for any $\xi \in \mathbb{R}$, $1 \leqslant p \leqslant 2$ and $r, \beta, \tau > 0$

$$\frac{\xi^2}{r^2} - \beta(\tau^2 + \xi^2)^{p/2} + \beta\tau^p \geqslant 0 \iff \beta\tau^p\left(1 + \frac{1}{\beta\tau^p}\frac{\xi^2}{r^2}\right) \geqslant \beta\tau^p\left(1 + \frac{\xi^2}{\tau^2}\right)^{p/2},$$

which holds true, using Bernoulli's inequality with exponent $p/2 \leqslant 1$, for any $0 < \beta \leqslant \frac{2\tau^{2-p}}{pr^2}$:

$$\left(1 + \frac{\xi^2}{\tau^2}\right)^{p/2} \leqslant 1 + \frac{p}{2}\frac{\xi^2}{\tau^2} \leqslant 1 + \frac{\xi^2}{\beta\tau^p r^2}.$$

By applying this observation componentwise with $r = \rho_j$ and $\tau = \gamma\rho_j$, we see that $f_{\rho,\beta,\gamma}$ is (globally) non-negative for any $0 < \beta \leqslant \min_{j=1,\dots,k}\frac{2\gamma^{2-p}}{p\rho_j^p} = \frac{2\gamma^{2-p}}{p\rho_1^p}$, proving (b) for any $1 \leqslant p \leqslant 2$ (for $\beta = 0$ the claim holds trivially). In the case $p > 2$, (b) follows from (4.1), since, for any $0 \leqslant \beta \leqslant \rho_1^{-2}$,

$$\sum_{j=1}^k \frac{x_j^2}{\rho_j^2} \geqslant \rho_1^{-2}\|x\|_2^2 \geqslant \beta\|x\|_p^2.$$

For (c), first consider the case $1 \leqslant p \leqslant 2$, for which the Hessian of $f_{\rho,\beta,\gamma}$ is diagonal. Hence $f_{\rho,\beta,\gamma}$ is convex if and only if all those diagonal entries,

$$\frac{\partial^2 f_{\rho,\beta,\gamma}}{\partial x_j^2}(x) = \frac{2}{\rho_j^2} - \beta p\frac{\gamma^2\rho_j^2 + (p-1)x_j^2}{(\gamma^2\rho_j^2 + x_j^2)^{2-p/2}}, \qquad j = 1, \dots, k,$$

are non-negative functions. Since, for $\tau > 0$, $\xi \in \mathbb{R}$ and $1 \leqslant p \leqslant 2$,

$$\frac{\tau^2 + (p-1)\xi^2}{(\tau^2 + \xi^2)^{2-p/2}} \leqslant \frac{\tau^2 + \xi^2}{(\tau^2 + \xi^2)^{2-p/2}} = \tau^{p-2}\frac{1 + \frac{\xi^2}{\tau^2}}{\left(1 + \frac{\xi^2}{\tau^2}\right)^{2-p/2}} \leqslant \tau^{p-2}, \tag{4.2}$$

$f_{\rho,\beta,\gamma}$ is convex for each $0 \leqslant \beta < \min_{j=1,\dots,k}\frac{2\gamma^{2-p}}{p\rho_j^p} = \frac{2\gamma^{2-p}}{p\rho_1^p}$ (by applying (4.2) componentwise with $\tau = \gamma\rho_j$, $j = 1, \dots, k$).

Now consider the case $2 < p < \infty$. The second-order partial derivatives of $L_{\rho,\gamma}$ for $x \neq 0$ are given by

$$\frac{\partial^2 L_{\rho,\gamma}}{\partial x_l \partial x_m}(x) = \begin{cases} \frac{2(p-1)|x_l|^{p-2}}{\|x\|_p^{p-2}} - \frac{2(p-2)|x_l|^{2p-2}}{\|x\|_p^{2p-2}} & \text{if } l = m, \\ -\frac{2(p-2)x_l x_m |x_l x_m|^{p-2}}{\|x\|_p^{2p-2}} & \text{if } l \neq m. \end{cases}$$

Hence, the Hessian of $f_{\rho,\beta,\gamma}$ for $x \neq 0$ can be written in the form

$$\nabla^2 f_{\rho,\beta,\gamma}(x) = \text{diag}\big((2\rho_j^{-2} - 2\beta(p-1)g_j(x))_{j=1,\dots,k}\big) + 2\beta(p-2)h(x)h(x)^{\mathsf{T}},$$

where $\mathrm{diag}(d_1,\ldots,d_k)$ denotes the $k \times k$ diagonal matrix with diagonal entries $d_1,\ldots,d_k$ and the functions $g_j\colon \mathbb{R}^k \setminus \{0\} \to \mathbb{R}, j = 1,\ldots,k$, and $h\colon \mathbb{R}^k \setminus \{0\} \to \mathbb{R}^k$ are given by

$$g_j(x) = \frac{|x_j|^{p-2}}{\|x\|_p^{p-2}}, \qquad h(x) = \left(\frac{x_j \cdot |x_j|^{p-2}}{\|x\|_p^{p-1}} \right)_{j=1,\ldots,k}.$$

Since $|g_j| \leqslant 1$, $\nabla^2 f_{\rho,\beta,\gamma}$ is symmetric and positive definite on the set $\mathbb{R}^k \setminus \{0\}$ for $0 \leqslant \beta < \frac{1}{(p-1)\rho_1^2}$. In order to prove convexity, we show that for any $x, y \in \mathbb{R}^k$ and $\lambda \in [0,1]$,

$$f_{\rho,\beta,\gamma}(\lambda x + (1-\lambda)y) \leqslant \lambda f_{\rho,\beta,\gamma}(x) + (1-\lambda)f_{\rho,\beta,\gamma}(y) \tag{4.3}$$

by considering the following three cases:

1. **case: $x,y \neq 0$ and the line through x and y does not touch the origin** $0 \in \mathbb{R}^k$. In this case, we can restrict the function $f_{\rho,\beta,\gamma}$ to an open half-space containing x and y, but not containing $0 \in \mathbb{R}^k$. On this convex set, $f_{\rho,\beta,\gamma}$ is twice continuously differentiable and positive definiteness of the Hessian $\nabla^2 f_{\rho,\beta,\gamma}$ proves convexity, in particular (4.3).
2. **case: $x,y \neq 0$ and the line through x and y contains the origin** $0 \in \mathbb{R}^k$. In this case, there exists $\lambda^\star \in (0,1)$ such that $\lambda^\star x + (1-\lambda^\star)y = 0$ and thereby $y = -\frac{\lambda^\star}{1-\lambda^\star}x$. It follows for each $\lambda \in [0,1]$ that

$$\lambda x + (1-\lambda)y = (\lambda - \lambda^\star)x + ((1-\lambda) - (1-\lambda^\star))y + 0$$

$$= (\lambda - \lambda^\star)(x-y) = \frac{\lambda - \lambda^\star}{1 - \lambda^\star}x.$$

Since $f_{\rho,\beta,\gamma}(tx) = t^2 f_{\rho,\beta,\gamma}(x)$ for each $t \in \mathbb{R}$,

$$g(\lambda) := f_{\rho,\beta,\gamma}(\lambda x + (1-\lambda)y) = f_{\rho,\beta,\gamma}\left(\frac{\lambda - \lambda^\star}{1 - \lambda^\star}x \right) = \left(\frac{\lambda - \lambda^\star}{1 - \lambda^\star} \right)^2 f_{\rho,\beta,\gamma}(x),$$

which is a quadratic function in λ with non-negative prefactor $f_{\rho,\beta,\gamma}(x) > 0$ (by (b)) and thereby convex. Therefore, we obtain (4.3) from

$$f_{\rho,\beta,\gamma}(\lambda x + (1-\lambda)y) = g(\lambda \cdot 1 + (1-\lambda) \cdot 0)$$

$$\leqslant \lambda g(1) + (1-\lambda)g(0)$$

$$= \lambda f_{\rho,\beta,\gamma}(x) + (1-\lambda)f_{\rho,\beta,\gamma}(y).$$

3. **case: $x \neq 0$ and $y = 0$.** In this case, (4.3) follows from the previous cases by continuity:

$$f_{\rho,\beta,\gamma}(\lambda x + (1-\lambda)y) = \lim_{t \searrow 0} f_{\rho,\beta,\gamma}(\lambda x + (1-\lambda)tx)$$

$$\leqslant \lim_{t \searrow 0} \lambda f_{\rho,\beta,\gamma}(x) + (1-\lambda)f_{\rho,\beta,\gamma}(tx)$$

$$= \lambda f_{\rho,\beta,\gamma}(x) + (1-\lambda)f_{\rho,\beta,\gamma}(y).$$

$\square$

Remark 4.7. Note that this bound on β is not optimal. For example, for $n = 2$, $p = 4$ and $\rho_1 = \rho_2 = 1$, we consider here $f_{\rho,\beta,\gamma}(x) = x^2 + y^2 - \beta\sqrt{x^4 + y^4}$. The lemma from above proves that this function is convex for $\beta < \frac{1}{3}$. In fact, it is convex even for $\beta < \sqrt{2}/3$ as can be shown by

more elementary methods (exclusive to this low-dimensional setting). Note that in this specific case $f_{\rho,\beta,\gamma}(x) \geqslant 0$ for $\beta \leqslant 1$.

Proposition 4.8. *Under assumptions 2.1 and 4.1 and using notation 4.3, for each $0 < \delta < 1$, each $k \in \mathbb{N} \cup \{0\}$, each $\gamma > 0$ and each $z \in X$,*

$$\mathfrak{R}_{\mu}^{\delta}(z,0) \leqslant \exp\left(- \frac{\gamma^{2-\alpha}}{4q\sigma_{k+1}^{\alpha}} \left((\|P_k z\|_p - \delta)^{\alpha} - \gamma^{\alpha} S^{\alpha} - \delta^{\alpha} \right) \right). \tag{4.4}$$

Proof. Let $\beta := \frac{\gamma^{2-\alpha}}{q\sigma_{k+1}^{\alpha}}$. Let $K \in \mathbb{N}$ and $\rho = (\sigma_{k+1}, \dots, \sigma_K)$.

Observe that the function $f \colon \mathbb{R}^k \to \mathbb{R}$ defined by

$$f(u) = \exp\left(- \tfrac{1}{4}|u|_{E^K}^2 - \tfrac{1}{4}|u|_{E^k}^2 - \tfrac{1}{4}(|u|_{E_k^K}^2 - \beta L_{\rho,\gamma}(P_k^K u)) \right)$$

is positive, symmetrical, integrable (since $f(u) \leqslant \exp(-\tfrac{1}{4}|u|_{E^K}^2)$ by proposition 4.6(b)) and log-concave (by proposition 4.6(c)). Hence, by propositions 4.6(a) and (c) and Anderson's inequality (theorem A.3),

$$\frac{\mu_K(B_\delta^K(z))}{\mu_K(B_\delta^K(0))} = \frac{\int_{B_\delta^K(z)} \exp\left(-\tfrac{1}{2}|u|_{E^K}^2 \right) \mathrm{d}u}{\int_{B_\delta^K(0)} \exp\left(-\tfrac{1}{2}|u|_{E^K}^2 \right) \mathrm{d}u}$$

$$\leqslant \frac{\sup_{v \in B_\delta^K(z)} \exp\left(-\tfrac{\beta}{4} L_{\rho,\gamma}(P_k^K v) \right)}{\inf_{v \in B_\delta^K(0)} \exp\left(-\tfrac{\beta}{4} L_{\rho,\gamma}(P_k^K v) \right)} \frac{\int_{B_\delta^K(z)} f(u)\mathrm{d}u}{\int_{B_\delta^K(0)} f(u)\mathrm{d}u}$$

$$\leqslant \exp\left(-\frac{\beta}{4} \left(\inf_{v \in B_\delta^K(z)} (\|P_k^K v\|_p^{\alpha} - \gamma^{\alpha}\|\rho\|_p^{\alpha}) - \sup_{v \in B_\delta^K(0)} \|P_k^K v\|_p^{\alpha} \right) \right)$$

$$\leqslant \exp\left(- \frac{\gamma^{2-\alpha}}{4q\sigma_{k+1}^{\alpha}} \left((\|P_k^K z\|_p - \delta)^{\alpha} - \gamma^{\alpha} S^{\alpha} - \delta^{\alpha} \right) \right).$$

For any $x \in X$, since $B_\delta(x) = \bigcap_{k \in \mathbb{N}} B_\delta^k(P^k x) \times \mathbb{R}^{\mathbb{N} \setminus \{1, \dots, k\}}$, the continuity of measures implies that $\mu_k(B_\delta^k(P^k x)) \to \mu(B_\delta(x))$. Therefore, taking the limit $K \to \infty$ proves the claim. $\qquad\square$

Corollary 4.9. *Under assumptions 2.1 and 4.1 the vanishing condition for unbounded sequences, condition 2.7 (C1), is satisfied.*

Proof. We use notation 4.3 throughout the proof. Let $(\delta_m)_{m \in \mathbb{N}}$ be a null sequence in $\mathbb{R}^+$ and $(x_m)_{m \in \mathbb{N}}$ be an unbounded sequence, i.e. there exists a subsequence $(x_{m_n})_{n \in \mathbb{N}}$ such that $\|x_{m_n}\|_p \to \infty$ as $n \to \infty$. Using notation 4.3 and proposition 4.8 with $\gamma = 1$ and $k = 0$ we obtain

$$\mathfrak{R}_{\mu}^{\delta_{m_n}(x_{m_n},0)} \leqslant \exp\left(- \frac{1}{4q\sigma_1^{\alpha}} \left((\|x_{m_n}\|_p - \delta_{m_n})^{\alpha} - S^{\alpha} - \delta_{m_n}^{\alpha} \right) \right) \xrightarrow[n \to \infty]{} 0,$$

proving the claim. $\qquad\square$

Corollary 4.10. *Under assumptions 2.1 and 4.1 the weakly convergent subsequence condition, condition 2.7 (C2), is satisfied.*

Proof. We use notation 4.3 throughout the proof. If $p > 1$, the statement follows directly from the reflexivity of $X = \ell^p$. Now let $p = 1$, let $(\delta_m)_{m \in \mathbb{N}}$ be a null sequence in $(0, 1)$ and $(x_m)_{m \in \mathbb{N}}$ be a bounded sequence in X satisfying, for some $K > 0$ and each $m \in \mathbb{N}$, $\mathfrak{R}_{\mu}^{\delta_m}(x_m, 0) \geqslant K$.

We first show that $(x_m)_{m \in \mathbb{N}}$ is equismall at infinity, i.e. for every $r > 0$ there exists $k \in \mathbb{N}$ such that, for each $m \in \mathbb{N}$, $\|P_k x_m\|_1 < r$. Assuming the contrary, there exists $r > 0$ such that, for any $k \in \mathbb{N}$, there exist $m_k \in \mathbb{N}$ such that $\|P_k x_{m_k}\|_1 \geqslant r$.

If the sequence $(m_k)_{k\in\mathbb{N}}$ was bounded by some $N \in \mathbb{N}$, then, using the fact that $\lim_{k\to\infty} \|P_k x\|_1 = 0$ for any (fixed) $x \in X$,

$$r \leqslant \limsup_{k\to\infty} \|P_k x_{m_k}\|_1 \leqslant \lim_{k\to\infty} \sup_{n=1,\dots,N} \|P_k x_n\|_1 = 0 < r.$$

Since this is a contradiction, $(m_k)_{k\in\mathbb{N}}$ is unbounded. Using $\sigma_k \searrow 0$ and $\delta_k \searrow 0$ as $k \to \infty$, this implies the existence of $k \in \mathbb{N}$ such that $\delta_{m_k} \leqslant r/8$ and

$$\exp\left(-\frac{r^2}{32\sigma_{k+1}\sum_{j\in\mathbb{N}}\sigma_j}\right) < K.$$

Using proposition 4.8 with $\gamma := \frac{r}{4\sum_{j\in\mathbb{N}}\sigma_j}$ we obtain

$$\mathfrak{R}_\mu^{\delta_{m_k}(x_{m_k},0)} \leqslant \exp\left(-\frac{\gamma}{4\sigma_{k+1}}\left(\|P_k x_{m_k}\|_1 - 2\delta_{m_k} - \gamma\sum_{j\in\mathbb{N}}\sigma_j\right)\right)$$

$$\leqslant \exp\left(-\frac{r}{16\sigma_{k+1}\sum_{j\in\mathbb{N}}\sigma_j}\left(r - \tfrac{r}{4} - \tfrac{r}{4}\right)\right)$$

$$\leqslant \exp\left(-\frac{r^2}{32\sigma_{k+1}\sum_{j\in\mathbb{N}}\sigma_j}\right)$$

$$< K,$$

contradicting the assumption $\mathfrak{R}_\mu^{\delta_m}(x_m,0) \geqslant K$ for each $m \in \mathbb{N}$.

Hence, $(x_m)_{m\in\mathbb{N}}$ is equismall at infinity and, combined with its boundedness, this implies the existence of a weakly convergent subsequence of $(x_m)_{m\in\mathbb{N}}$ by (Trèves 1967, theorem 44.2). $\qquad\square$

Corollary 4.11. *Under assumptions 2.1 and 4.1 the vanishing condition for weakly, but not strongly convergent sequences, condition 2.7 (C4), is satisfied.*

Proof. We use notation 4.3 throughout the proof. Let $(\delta_m)_{m\in\mathbb{N}}$ be a null sequence in $\mathbb{R}^+$ and $(x_m)_{m\in\mathbb{N}}$ be a weakly, but not strongly convergent sequence in X with weak limit $\bar{z} \in E$,

Step 1: There exists a $c > 0$ and $k_0 \in \mathbb{N}$ such that, for any $k \geqslant k_0$,

$$\limsup_{m\to\infty} \|P_k x_m\|_X > c.$$

There exists $A > 0$ such that $\limsup_{m\to\infty} \|x_m - \bar{z}\|_X > A$ (otherwise the convergence would be strong). Let $c := \frac{A}{2}$. Since $\bar{z} \in E$, we have $|P_k \bar{z}|_E \to 0$ as $k \to \infty$ by lemma 4.4 and therefore $\|P_k \bar{z}\|_X \to 0$ as $k \to \infty$ by continuous embedding $E \subset X$ (Bogachev 1998, proposition 2.4.6). Hence, there exists $k_0 \in \mathbb{N}$ such that, for each $k \geqslant k_0$, $\|P_k \bar{z}\|_X < c$. Let $k \geqslant k_0$ and assume the contrapositive, i.e. $\limsup_{m\to\infty} \|P_k x_m\|_X \leqslant c$. But then, since weak convergence implies componentwise convergence,

$$2c = A < \limsup_{m\to\infty}\|x_m - \bar{z}\|_X = \limsup_{m\to\infty}\|P^k(x_m - \bar{z}) + P_k x_m - P_k\bar{z}\|_X$$

$$\leqslant \underbrace{\limsup_{m\to\infty}\|P^k(x_m-\bar{z})\|_X}_{=0\text{ by weak conv.}} + \underbrace{\limsup_{m\to\infty}\|P_k x_m\|_X}_{\leqslant c\text{ by assumption}} + \underbrace{\|P_k\bar{z}\|_X}_{<c\text{ since }k\geqslant k_0}$$

$$< 2c,$$

which is a contradiction, proving the claim.

Step 2: For each $0 < \varepsilon < 1$, $\liminf_{m \to \infty} \mathfrak{R}_\mu^{\delta_m}(x_m, 0) < \varepsilon$.
Let $0 < \varepsilon < 1$, $\delta_0 := \frac{c}{4}$, $\gamma := \frac{c}{4S}$ and $k \geqslant k_0$ such that

$$\sigma_{k+1} < \left(\frac{c^2}{4^{4-\alpha} S^{2-\alpha} q(-\log \varepsilon)} \right)^{1/\alpha}.$$

Let $m_0 \in \mathbb{N}$. Using Step 1, there exists $m \geqslant m_0$ such that $\delta_m < \delta_0 = \frac{c}{4}$ and $\|P_k x_m\|_X > c$. Since $\frac{3^\alpha - 2}{4^\alpha} \geqslant \frac{1}{4}$ for $1 \leqslant \alpha \leqslant 2$, and by setting $\gamma = \frac{c}{4S}$, proposition 4.8 implies

$$\mathfrak{R}_\mu^{\delta_m}(x_m, 0) \leqslant \exp \left(-\frac{\gamma^{2-\alpha}}{4q\sigma_{k+1}^\alpha} \left((\|P_k x_m\|_p - \delta_m)^\alpha - \gamma^\alpha S^\alpha - \delta_m^\alpha \right) \right)$$

$$\leqslant \exp \left(-\frac{\left(\frac{c}{4S} \right)^{2-\alpha}}{4q\sigma_{k+1}^\alpha} \left(\left(\frac{3c}{4} \right)^\alpha - \left(\frac{c}{4} \right)^\alpha - \left(\frac{c}{4} \right)^\alpha \right) \right)$$

$$\leqslant \exp \left(-\frac{c^2}{4^{4-\alpha} S^{2-\alpha} q\sigma_{k+1}^\alpha} \right)$$

$$< \varepsilon.$$

$\square$

Proof of theorem 2.5. By lemma 4.5, corollaries 4.9–4.11, condition 2.7 (C1)–(C4) are fulfilled and all statements follow from theorem 2.8. $\square$

5. Conclusion

We proved the existence of MAP estimators in the context of a Bayesian inverse problem for parameters in a separable Banach space X, where X is either a Hilbert space or $X = \ell^p$, $p \in [1, \infty)$, with a diagonal Gaussian prior. The Hilbert space case had been proven before by (Dashti *et al* 2013, Kretschmann 2019), however, they did not show the existence of the central object in their proofs, namely the δ-ball maximizers $z^\delta = \operatorname{argmax}_{z \in X} \mu^y(B_\delta(z))$. We fixed this gap by working with an AMF $(\zeta^\delta)_{\delta > 0} \subset X$ defined by definition 1.7 and strongly simplified their proof by employing (Da Prato and Zabczyk 2002, proposition 1.3.11), restated in proposition 3.2. We decided to present this elegant and simple proof even though the Hilbert space case can be understood as a special case of $X = \ell^p$ for $p = 2$. The case $p \neq 2$, on the other hand, turned out to require novel techniques to prove the corresponding results. The crucial mathematical argument in this case relies on a convexification of the difference $|\cdot|_E^2 - \beta\|\cdot\|_X^2$ (proposition 4.6). This allows to extract a suitable 'rate of contraction' such that the ratio $\mathfrak{R}_\mu^\delta(z, 0)$ can be bounded for any fixed $\delta > 0$ by a function decaying exponentially in $\|z\|_X$ (proposition 4.8).

We have also outlined a general proof strategy in section 2 how similar results (i.e. conjecture 2.3) can be obtained for further separable Banach spaces. For this purpose, we filtered out four crucial conditions, namely condition 2.7 (C1)–(C4), which need to be proven in the Banach space of interest, and then the corresponding result follows almost immediately from theorem 2.8.

Note that our results rely strongly on the characteristics of the ℓ^p norm and the diagonal structure of the covariance matrix of the Gaussian measure. We suspect that the generalization to Gaussian measures on arbitrary separable Banach spaces requires deeper insight into the compatibility between the ambient space's geometry and the Cameron–Martin norm. We hope that our theorem 2.8 paves the way for future research in this direction.

Data availability statement

No new data were created or analysed in this study.

Acknowledgments

Appendix. Gaussian measures in Banach spaces

In notation, we will mainly follow (Bogachev 1998). The continuous (or topological) dual space of X is denoted by $X^\star$, while X' denotes its algebraic dual. In some cases, we will assume that X is a Hilbert space, in which case we write $X = \mathscr{H}$ for clarity. The object μ will always be a centered Gaussian measure on X (or $\mathscr{H}$). We denote the Cameron–Martin space by $(E, \langle \cdot, \cdot \rangle_E)$, where we write the Cameron–Martin norm with single bars in order to differentiate it from the ambient space norm: $|u|_E := \sqrt{\langle u, u \rangle_E}$.

It turns out that the extension of the covariance operator

$$R_\mu \colon X^\star \to (X^\star)', \qquad (R_\mu f)(g) := \langle f, g \rangle_{L^2(\mu)}$$

to the reproducing kernel Hilbert space (RKHS) $X_\mu^\star := \overline{X^\star}^{L^2(X,\mu)}$ of μ satisfies $R_\mu(X_\mu^\star) = E$ (Bogachev 1998, theorem 3.2.3), where E is viewed as a subspace of $(X^\star)'$. In addition, $R_\mu \colon (X_\mu^\star, \langle \cdot, \cdot \rangle_{L^2(\mu)}) \to (E, \langle \cdot, \cdot \rangle_E)$ is an isometric isomorphism (Bogachev 1998, p 60) and satisfies the *reproducing property*

$$f(h) = \langle R_\mu f, h \rangle_E, \qquad f \in X_\mu^\star, \quad h \in E, \tag{A.1}$$

which follows from the above and from treating $h = R_\mu g$ (for some $g \in X_\mu^\star$) as an element of $(X^\star)'$:

$$f(h) = f(R_\mu g) = (R_\mu g)(f) = \langle f, g \rangle_{L^2(\mu)} = \langle R_\mu f, R_\mu g \rangle_E = \langle R_\mu f, h \rangle_E.$$

Remark A.1. In the special case where the measure is defined on a Hilbert space $\mathscr{H}$, the covariance operator R_μ takes the form of a self-adjoint, non-negative trace-class operator: $R_\mu = Q$ where

$$Q \colon Q^{-1/2}(X) = X_\mu^\star \to E = Q^{1/2}X.$$

In addition, the CM inner product and norm take the form

$$\langle u, v \rangle_E = \langle Q^{-1/2}u, Q^{-1/2}v \rangle_\mathscr{H}, \qquad |u|_E = \|Q^{-1/2}u\|_\mathscr{H}. \tag{A.2}$$

A result we are going to use in this context is the following technical lemma:

Lemma A.2. *Let X be a separable Banach space and μ a centered Gaussian measure on X, $\bar{z} \in E$ and $x^\delta \rightharpoonup \bar{z}$ weakly in X. Then*

$$\limsup_{\delta \to 0} \mathfrak{R}_\mu^\delta(x^\delta, \bar{z}) \leqslant 1.$$

Proof. For any $\hat{h} \in X^\star$, the Cameron–Martin formula (Bogachev 1998, corollary 2.4.3) implies

$$
\mu(B_\delta(x^\delta)) = \int_{B_\delta(x^\delta)} \mathrm{d}\mu = \int_{B_\delta(x^\delta - R_\mu \hat{h})} \exp\left(-\tfrac{1}{2}|R_\mu \hat{h}|_E^2 - \hat{h}(u)\right) \mathrm{d}\mu(u)
$$

$$
\leqslant \mu(B_\delta(x^\delta - R_\mu \hat{h})) \exp\left(-\tfrac{1}{2}|R_\mu \hat{h}|_E^2\right) \sup_{u \in B_\delta(x^\delta - R_\mu \hat{h})} e^{-\hat{h}(u)}
$$

$$
\leqslant \mu(B_\delta(0)) \exp\left(-\tfrac{1}{2}|R_\mu \hat{h}|_E^2 - \hat{h}(x^\delta - R_\mu \hat{h})\right) \sup_{u \in B_\delta(0)} e^{-\hat{h}(u)}, \quad \text{(A.3)}
$$

where we used Anderson's inequality (theorem A.4) in the last step. Since

$$
\int_{B_\delta(0)} \exp(-(R_\mu^{-1} \bar{z})(u)) \, \mathrm{d}\mu(u) = \int_{B_\delta(0)} \exp((R_\mu^{-1} \bar{z})(u)) \, \mathrm{d}\mu(u)
$$

due to symmetry of the set $B_\delta(0)$, another application of the Cameron–Martin theorem yields

$$
\mu(B_\delta(\bar{z})) = \exp\left(-\tfrac{1}{2}|\bar{z}|_E^2\right) \int_{B_\delta(0)} \exp(-(R_\mu^{-1} \bar{z})(u)) \, \mathrm{d}\mu(u)
$$

$$
= \exp\left(-\tfrac{1}{2}|\bar{z}|_E^2\right) \int_{B_\delta(0)} \frac{\exp((R_\mu^{-1} \bar{z})(u)) + \exp(-(R_\mu^{-1} \bar{z})(u))}{2} \, \mathrm{d}\mu(u)
$$

$$
\geqslant \exp\left(-\tfrac{1}{2}|\bar{z}|_E^2\right) \mu(B_\delta(0)), \quad \text{(A.4)}
$$

where we used the inequality $a + a^{-1} \geqslant 2$ for any $a > 0$ (alternatively, (A.4) can be proven via Jensen's inequality). Since $x^\delta \to \bar{z}$ weakly in X, it follows from (A.3) and (A.4) that, for any $\hat{h} \in X^\star$,

$$
\limsup_{\delta \searrow 0} \mathfrak{R}_\mu^\delta(x^\delta, \bar{z}) \leqslant \limsup_{\delta \searrow 0} \exp\left(\tfrac{1}{2}|\bar{z}|_E^2 - \tfrac{1}{2}|R_\mu \hat{h}|_E^2 - \hat{h}(x^\delta - R_\mu \hat{h})\right) \sup_{u \in B_\delta(0)} e^{-\hat{h}(u)}
$$

$$
\leqslant \exp\left(\tfrac{1}{2}|\bar{z}|_E^2 - \tfrac{1}{2}|R_\mu \hat{h}|_E^2 - \hat{h}(\bar{z} - R_\mu \hat{h})\right)
$$

$$
= \exp\left(\tfrac{1}{2}|\bar{z}|_E^2 - \tfrac{1}{2}|R_\mu \hat{h}|_E^2 - \langle R_\mu \hat{h}, \bar{z} - R_\mu \hat{h}\rangle_E\right)
$$

$$
= \exp\left(\tfrac{1}{2}|\bar{z}|_E^2 - \tfrac{1}{2}|R_\mu \hat{h}|_E^2 - \langle \bar{z}, \bar{z} - R_\mu \hat{h}\rangle_E + |R_\mu \hat{h} - \bar{z}|_E^2\right),
$$

where we used the reproducing property (A.1). Choosing a sequence $(\hat{h}_n)_{n \in \mathbb{N}}$ in $X^\star$ such that $R_\mu \hat{h}_n \to \bar{z}$ strongly in E (this is possible by density of $X^\star$ in $R_\mu^{-1} E$), replacing $\hat{h}$ by $\hat{h}_n$ in the above inequality and taking the limit $n \to \infty$ proves the claim. $\qquad \square$

Theorem A.3 (Anderson's inequality, version 1; Bogachev 2007, theorem 3.10.25). *Let A be a bounded centrally symmetric convex set in $\mathbb{R}^n$, $n \in \mathbb{N}$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be*

- *non-negative and locally integrable,*
- *symmetrical, i.e. $f(-x) = f(x)$ for each $x \in \mathbb{R}^n$, and*
- *unimodal, i.e. the sets $\{f \geqslant c\}$ are convex for all $c > 0$.*

26

Then, for every $h \in \mathbb{R}^n$ and every $t \in [0, 1]$, one has

$$\int_A f(x + th)\, dx \geqslant \int_A f(x + h)\, dx.$$

In particular, for every $z \in \mathbb{R}^n$, $\int_{z+A} f(x)\, dx \leqslant \int_A f(x)\, dx$.

Theorem A.4 (Anderson's inequality, version 2; Bogachev 1998, corollary 4.2.3). *Let γ be a centered Gaussian measure on a Banach space X. Let A be a centrally symmetric convex set. Then for any $a \in X$, we have that $\gamma(A + a) \leqslant \gamma(A)$.*

ORCID iDs

Ilja Klebanov ⓘ https://orcid.org/0000-0002-1911-4281
Philipp Wacker ⓘ https://orcid.org/0000-0001-8718-4313

References

Agapiou S, Burger M, Dashti M and Helin T 2018 Sparsity-promoting and edge-preserving maximum *a posteriori* estimators in non-parametric Bayesian inverse problems *Inverse Problems* **34** 045002

Ayanbayev B, Klebanov I, Lie H C and Sullivan T J 2021a Γ-convergence of Onsager–Machlup functionals: I. With applications to maximum *a posteriori* estimation in Bayesian inverse problems *Inverse Problems* **38** 025005

Ayanbayev B, Klebanov I, Lie H C and Sullivan T J 2021b Γ-convergence of Onsager–Machlup functionals: II. Infinite product measures on Banach spaces *Inverse Problems* **38** 025006

Baker C R 1973 Joint measures and cross-covariance operators *Trans. Am. Math. Soc.* **186** 273–89

Bogachev V I 1998 *Gaussian Measures* (*Mathematical Surveys and Monographs* vol 62) (Providence, RI: American Mathematical Society)

Bogachev V I 2007 *Measure Theory* vol I, II (Berlin: Springer)

Clason C, Helin T, Kretschmann R and Piiroinen P 2019 Generalized modes in Bayesian inverse problems *SIAM/ASA J. Uncertain. Quantif.* **7** 652–84

Da Prato G and Zabczyk J 2002 *Second Order Partial Differential Equations in Hilbert Spaces* (*London Mathematical Society Lecture Note Series* vol 293) (Cambridge: Cambridge University Press)

Da Prato G and Zabczyk J 2014 *Stochastic Equations in Infinite Dimensions* (*Encyclopedia of Mathematics and its Applications* vol 152) 2nd edn (Cambridge: Cambridge University Press)

Dashti M, Law K J H, Stuart A M and Voss J 2013 MAP estimators and their consistency in Bayesian nonparametric inverse problems *Inverse Problems* **29** 095017

Dürr D and Bach A 1978 The Onsager–Machlup function as Lagrangian for the most probable path of a diffusion process *Comm. Math. Phys.* **60** 153–70

Helin T and Burger M 2015 Maximum *a posteriori* probability estimates in infinite-dimensional Bayesian inverse problems *Inverse Problems* **31** 085009

Kretschmann R 2019 Nonparametric Bayesian inverse problems with Laplacian noise *PhD Thesis* Universität Duisburg-Essen

Lambley H and Sullivan T J 2022 An order-theoretic perspective on modes and maximum *a posteriori* estimation in Bayesian inverse problems (arXiv:2209.11517)

Lie H C and Sullivan T J 2018 Equivalence of weak and strong modes of measures on topological vector spaces *Inverse Problems* **34** 115013

Sprungk B 2017 Numerical methods for Bayesian inference in Hilbert spaces *PhD Thesis* Technische Universität Chemnitz

Stuart A M 2010 Inverse problems: a Bayesian perspective *Acta Numer.* **19** 451–559

Trèves F 1967 *Topological Vector Spaces, Distributions and Kernels* (New York: Academic)